DOCUMENT RESUME

ED 096 230	SO 007 774

AUTHOR	Krueger, Reynold J.
TITLE	An Investigation of Personality and Music Teaching Success. Final Report.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Regional Research Program.
BUREAU NO	BR-2-E-015
PUB DATE	May 74
GRANT	OEG-5-72-0019(509)
NOTE	435p.
EDRS PRICE	MF-$0.75 HC-$21.00 PLUS POSTAGE
DESCRIPTORS	College Students; Educational Research; Education Majors; Effective Teaching; *Music Education; *Music Teachers; *Occupational Guidance; Performance Factors; *Personality Assessment; *Success Factors; Teacher Motivation; Teacher Rating; Vocational Counseling

ABSTRACT

The purpose of the study was to investigate the relationships between personality and motivational variables and music teaching success criteria. The success criteria investigated were class mean residual gain scores of pupils and teacher ratings by pupils, building principals, supervisors, peers, and teachers themselves. The problems considered in the study were the relationships among the success criteria; the relationships and interrelationships of the personality variables and motivational variables; personality profiles for successful and unsuccessful male and female music educators; and the implications of the use of personality measures for counseling and advising undergraduate music education students. The subjects of the study were 209 music educators and 73 music student teachers. The report of the study includes chapters stating the problem; surveying the related literature; describing procedures and analysis of data; summarizing the findings, which strongly related personality and motivational factors to music teaching success; and making recommendations. A bibliography and various appendixes of participating school districts, correspondence, and forms used in the study complete the report. (JH)

ED 096230

BEST COPY AVAILABLE

Final Report

Project No. 2-E-015
Grant No. OEG-5-72-0019 (509)

Reynold J. Krueger
The Board of Trustees of
The University of Illinois
Urbana, Illinois 61820

U S DEPARTMENT OF HEALTH
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

AN INVESTIGATION OF PERSONALITY AND MUSIC TEACHING SUCCESS

May 1974

SP 007 774

U.S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education

(Regional Research Program)

BEST COPY AVAILABLE

ABSTRACT

The purpose of the study was to investigate the relationship between personality and motivational variables and music teaching success. Six criteria were selected for investigation: 1. class mean residual gain scores of pupils (G), 2. teacher ratings by pupils, 3. teacher ratings by principals, 4. teacher ratings by supervisors, 5. teacher ratings by peers, and 6. teacher self ratings.

The 5 ratings were factor analyzed from which the first principal component was used to generate a single factor rating score (R) for each of the 209 music educators and 73 music student teachers. Four criterion variables were constructed: gain (G), rating (R), gain plus rating (G+R), and gain minus rating (G-R).

Multiple regression and discriminant function analyses investigated the relationships. Significant relationships beyond the .05 level were found and cross validated between the motivational variables measured by the Motivation Analysis Test and the G and G-R criteria. The relationships between the 16PF test and the 4 criterion variables were significant but did not cross validate. There were significant profile differences beyond the .05 level between successful and unsuccessful groups of male and female music educators for both personality trait and motivational variables when stratified for both G and R criteria.

BEST COPY AVAILABLE

Final Report

Project No. 2-E-015
Grant No. OEG-5-72-0019 (509)

AN INVESTIGATION OF PERSONALITY AND MUSIC TEACHING SUCCESS

Reynold J. Krueger

The University of Illinois

Urbana, Illinois 61820

May 1974

The research reported herein was performed pursuant to a grant with the Office of Education, U.S. Department of Health, Education and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education

(Regional Research Program)

BEST COPY AVAILABLE

ACKNOWLEDGMENTS

The writer sincerely appreciates the encouragement and assistance offered by Dr. Richard J. Colwell, Professor of Music and Education, who gave freely of his time and energy. His influence upon the writer as teacher, adviser, committee chairman, and friend is profound.

Warm thanks are offered the scholars comprising the writer's committee who were generous with their time and comments: Dr. Lawrence M. Aleamoni, Professor of Educational Psychology; Dr. Charles Leonhard, Professor of Music; and Dr. Maurice M. Tatsuoka, Professor of Psychology and Educational Psychology. Their guidance is cherished.

This study could not have been conducted without the active support of music educators, administrators, supervisors, and consultants who elicited the participation of subjects, coordinated the data gathering procedures, and assisted in the interpretation of the findings. Their contributions were many:

Mrs. Portia Bailey, Chicago, Illinois, Public Schools
Mrs. Lois Best, Tacoma, Washington, Public Schools
Dr. James M. Brinkman, Eastern Illinois University
Mr. Delmar E. Burgin, University of Illinois
Dr. Howard G. Doolin, Dade County Public Schools, Miami, Florida
Dr. Walter B. Duda, Cicero, Illinois, Public Schools
Mr. Marlowe Earle, Fountain Valley, California, Public Schools
Mr. Alan L. Farrell, Paducah, Kentucky, Public Schools
Dr. Kenneth Geoffroy, South Bend, Indiana, Public Schools
Dr. A. Oren Gould, Western Illinois University
Mr. Thaddeaus Grimes, Gainesville, Florida, Public Schools
Miss Edna Hehn, Oakland, California, Public Schools
Mr. Robert G. Howell, Joliet, Illinois, Public Schools
Mr. Eugene Johnson, Quincy, Illinois, Public Schools

BEST COPY AVAILABLE

Mrs. Marilyn Ivey Jokisch. Villa Park, Illinois, Public Schools

Mr. Delwen B. Jones, Tacoma, Washington, Public Schools

Mr. Robert Jorgensen, Urbana, Illinois, Public Schools

Prof. Dale Kimpton, University of Illinois

Mrs. Sunny Lawrence, Villa Park, Illinois, Public Schools

Dean Fred C. Mayer, Oklahoma City University

Mr. Robert McSpadden, Cincinnati, Ohio, Public Schools

Dr. Rees G. Olson, Buena Park, California, Public Schools

Mrs. Ruth Parson, Abilene, Texas, Public Schools

Mr. Orlo Rees, Garden Grove, California, Public Schools

Prof. Paul Rosene, Illinois State University

Mr. Verrollton C. Shaul, Champaign, Illinois, Public Schools

Dr. James Sorensen, Elmhurst College

Dr. Robert E. Thomas, University of Illinois

Dr. Russell L. Tiede, Millikin University

Dr. Loren Waa, Fargo, North Dakota, Public Schools

Dr. J. William Worrel, Cincinnati, Ohio, Public Schools

Thanks go to Dr. Roger Edwards, CEMREL, and Dr. Sam Krug, IPAT, for their wisdom, advice, encouragement, and statistical assistance. The writer is grateful to his secretary, Mrs. Edward Doczy, whose administrative, stenographic, and communicative skills contributed innumerably to the success of this investigation. Furthermore, the writer is indebted to Mrs. Russell Tiede for superbly typing the final report.

Special recognition and loving thanks are extended the writer's parents, Mr. and Mrs. Jack E. Krueger, whose moral and financial support assisted him over many obstacles. While it is not usually proper to mention a finding in

BEST COPY AVAILABLE

the "Acknowledgments" section: the writer reports that had he not undertaken this investigation, he would not have found nor married his beautiful wife Barbara, a participating music educator.

Finally, it would be inappropriate not to thank the music educators and music student teachers who were the subjects in this study. Their contributions are greatly appreciated.

BEST COPY AVAILABLE

TABLE OF CONTENTS

CHAPTER Page

I. THE STATEMENT OF THE PROBLEM....1

Criteria of Music Teaching Success....14
Personality and Motivational Variables....30
The 16 Primary Source Trait Factors....36

II. A SURVEY OF RELATED LITERATURE....60

Section One: Predictive Measurement of Success....64
Section Two: Teacher Accountability Criteria....106
Section Three: The Relationships of Personality to Music Teaching Success....146

III. PROCEDURES....179

Data Gathering....180
Scoring and Tabulation of Data....188
Data Analyses....195

IV. ANALYSIS OF DATA AND FINDINGS....204

Sample Attrition....204
The Pupil Growth Criterion of Music Teaching Success....218
Ratings of Music Teaching Success....244
Relationships Among Accountability Criteria....268
The Relationships and Interrelationships Among Personality, Motivational, and Criterion Variables....275
Male Instrumental....298
Male Choral and Female Instrumental....308
Female Choral....310
Personality and Motivation Profiles....323
Advising and Counseling Music Education Majors....333

V. SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS....342

Summary of the Findings....350
Conclusions....368
Recommendations....373

BIBLIOGRAPHY....377

APPENDIX

A. LIST OF PARTICIPATING SCHOOL DISTRICTS....388
B. LETTERS....391
C. FORMS USED IN THE STUDY....411

VITA....424

LIST OF TABLES

BEST COPY AVAILABLE

Table Page

1. 16PF Primary Factors.....34
2. MAT Factors.....49
3. MAT Pretesting Times.....184
4. Music Educator Attrition.....209
5. Music Educator Attrition - Paid and Unpaid.....210
6. Music Educator Sample After Attrition.....210
7. Socioeconomic School Category Distribution.....211
8. Female Choral In Study.....213
9. Female Choral Out Of Study.....213
10. Female Instrumental In Study.....214
11. Female Instrumental Out Of Study.....214
12. Male Choral In Study.....215
13. Male Choral Out Of Study.....215
14. Male Instrumental In The Study.....216
15. Male Instrumental Out Of Study.....216
16. Music Student Teacher Sample.....217
17. Student Teacher Sample Attrition.....218
18. Class Mean Residual Gain Score Analysis.....220
19. ANOVA Class Mean Residual Gain Scores.....221
20. t-Test Analysis Among Pretesting Groups.....221
21. Music Achievement Test Subtests.....223
22. Female Choral Gain Success.....226
23. Male Instrumental Gain Scores.....229
24. Male Choral Gain Scores.....231
25. Female Instrumental Gain Scores.....232
26. Gain Score Distribution by Groups.....235

BEST COPY AVAILABLE

Table Page

27. Gain Score Within Group Percentages.......................................235

28. Gain Score Distribution - Socioeconomic...................................236

29. Socioeconomic Gain Score Distribution Within Groups..................237

30. Gain Score Distribution By Grade Level.....................................240

31. Music Achievement Test Distribution...242

32. Music Educator Rating Data...245

33. Music Educator Rating Correlation Matrix....................................246

34. Pupil Rating Predictions..247

35. Principal Rating Predictions...247

36. Music Supervision Rating Predictions..247

37. Peer Teacher Rating Predictions...247

38. Pupil and Supervisor Predictions..248

39. Pupil and Peer Predictions..248

40. Pupil and Self Predictions...248

41. Principal and Supervisor Predictions..249

42. Supervisor and Peer Predictions..249

43. Peer and Self Predictions..249

44. Student Teacher Rating Mean Scores...250

45. Student Teachef Rating Correlations...251

46. Student Teacher Missing Data Predictions.......................................252

47. Pupil Predictions..253

48. Cooperating Teacher Predictions..253

49. Supervising Teacher Predictions...254

50. Peer Predictions...254

51. Pupil and Cooperating Teacher Predictions......................................255

52. Pupil and Supervising Teacher Predictions.......................................255

53. Pupil and Self Predictions...256

BEST COPY AVAILABLE

Table Page

54. Cooperating Teacher and Supervisor Predictions....................256

55. Cooperating Teacher and Self Predictions....................257

56. Supervisor and Peer Predictions....................257

57. Supervisor and Self Predictions....................258

58. Peer and Self Predictions....................258

59. Ratings ANOVA....................260

60. Music Educator Rating Correlations....................261

61. Factor Analysis of 5 Ratings....................262

62. Student Teacher Rating Correlations....................265

63. Student Teacher Factor Analysis....................266

64. Gain, Rating, and Sex Correlations....................270

65. Criterion Data Intercorrelations....................272

66. 16PF and Criteria Correlations....................276

67. MAT and Criteria Correlations....................278

68. 16PF Variables....................283

69. 16PF Regression Analysis - Gain....................285

70. 16PF Regression Analysis - Rating....................285

71. 16PF Regression Analysis - Gain + Rating....................287

72. 16PF Regression Analysis - Gain - Rating....................287

73. Motivation Analysis Test Variables....................289

74. MAT Regression Analysis - Gain....................291

75. MAT Regression Analysis - Rating....................293

76. MAT Regression Analysis - Gain + Rating....................294

77. MAT Regression Analysis - Gain - Rating....................294

78. 16PF Regression Analysis - Rating - Male Instrumental....................299

79. 16PF Regression Analysis - Gain - Male Instrumental....................301

80. 16PF Regression Analysis - Gain + Rating - Male Instrumental....................302

BEST COPY AVAILABLE

Table Page

81. 16PF Regression Analysis - Gain - Rating - Male Instrumental..........303

82. MAT Regression Analysis - Rating - Male Instrumental..................304

83. MAT Regression Analysis - Gain - Male Instrumental....................306

84. MAT Regression Analysis - Gain + Rating - Male Instrumental...........307

85. MAT Regression Analysis - Gain - Rating - Male Instrumental...........309

86. 16PF Regression Analysis - Rating - Female Choral.....................311

87. 16PF Regression Analysis - Gain - Female Choral.......................312

88. 16PF Regression Analysis - Gain + Rating - Female Choral..............314

89. 16PF Regression Analysis - Gain - Rating - Female Choral..............315

90. MAT Regression Analysis - Rating - Female Choral......................317

91. MAT Regression Analysis - Gain - Female Choral........................318

92. MAT Regression Analysis - Gain + Rating - Female Choral...............319

93. MAT Regression Analysis - Gain - Rating - Female Choral...............320

94. Comparisons of Successful and Unsuccessful Music Educator Profiles for Rating and Gain Criteria Separately by Sex for the 16PF Test......326

95. Comparisons of Successful and Unsuccessful Music Educator Profiles for Rating and Gain Criteria Separately by Sex for the MAT Test........326

96. Comparisons of Male and Female Music Educator Profiles for Level of Success Separately by Rating and Gain Criteria, 16PF Test..........330

97. Comparisons of Male and Female Music Educator Profile for Level of Success Separately by Rating and Gain Criteria, MAT Test...........330

BEST COPY AVAILABLE

CHAPTER I

THE STATEMENT OF THE PROBLEM

Introduction

Improving the education profession has long been a concern of educators, psychologists, parents, and students. Numerous approaches and theories have been formulated by concerned persons for changing the ways in which learning takes place in the schools. Crucial to any educational endeavor is the teacher because how he perceives his profession and role in life affect his behavior and student learning.

> To find out what one is fitted to do and to secure an opportunity to do it is the key to happiness. Nothing is more tragic than failure to discover one's true business in life, or to find that one has drifted or been forced by circumstances into an uncongenial calling (Dewey 1916:308).

Teachers achieve varying degrees of success. Some are unhappy, frustrated, and disappointed while others thrive on the satisfactions received from daily contacts with students. Personality and motivational characteristics appear to be predominant factors in deciding whether one chooses teaching for a career or finds fulfillment, satisfaction, and success in other professions or occupations. David Ausubel states the case most succinctly in discussing the teacher's personal commitment to the intellectual development of students: "It determines in large measure whether he will expend the necessary effort to teach for real gains in the intellectual growth of pupils, or will merely go through the formal motions of teaching" (Ausubel 1968:412). Differences in teaching

BEST COPY AVAILABLE

effectiveness and job satisfaction may be related to teacher personality dimensions.

> The remaking of American public education requires, and indeed will not be possible without, fundamental changes in the education of teachers -- without, in a sense, the creation of a new breed of teacher-educator, educated to self scrutiny and to serious thought about purpose (Silberman 1970:374).

This call for a new breed of teacher implies that new procedures are needed for identifying effective teachers as well as for training prospective teachers. This has been recognized by the Music Educators National Conference in a recent publication highlighting the results of various task forces.

> A new type of teacher education in music is needed in order to provide "music educators who are competent, flexible, curious, and prepared to survive and flourish in a world of change" . . . clearly, a different kind of preparation is needed (MENC 1972:4).

Today, there is a controversy about educational evaluation and teacher accountability. Various programs have been offered and evaluation models derived such as PPBS, PERT, National Assessment, formative and summative -- process and product evaluations, responsive models, and a plethora of techniques for producing change in assisting pupils' learnings. Numerous workshops, clinics, and meetings include evaluation sessions and nary is there a professional meeting or convention today which does not in some context highlight teacher accountability and suggest means for achieving and evaluating pupil and teacher success. Perhaps James Cass summarized the teacher accountability dilemma best:

> As we focus increasingly on pupil performance as a measure of teacher effectiveness, however, it would be easy to forget the complexity of the learning process -- that individual children are very different, that they learn things at different rates, and that even the same child learns at a different rate at different times. If, therefore, the laudable effort to improve classroom practice by assessing teacher and school effectiveness merely results in placing more intense and sophisticated pressure on the children to perform, the very principle will be denied in practice, for if the concept means anything, it is that the ultimate accountability must be to the children (Cass 1971:41).

In 1922, William A. McCall enumerated 10 fundamental assumptions underlying scientific procedures for rating and evaluating teachers and these appear

BEST COPY AVAILABLE

to be valid methodological principles today:

> 1. The pupil is the center of gravity or sun of the educational system. Teachers are satellites of this sun and supervisors are moons of the satellites.
>
> 2. All the paraphernalia of education exist for just one purpose, to make desirable changes in pupils.
>
> 3. The worth of these paraphernalia can be measured in just one way, by determining how many desirable changes they make in pupils.
>
> 4. Hence, the only just basis for selecting and promoting teachers is the changes made in pupils.
>
> 5. Teachers are at present selected and promoted primarily on the basis of their attributes, such as intelligence, personality, physical appearance, voice, ability in penmanship, and the like.
>
> 6. No one has demonstrated just what causal relationship, if any, exists between possession of these attributes and desirable changes in pupils . . .
>
> 7. Scientific measurement itself is fair only when we measure the amount of desirable change produced in pupils by a given teacher. The measurement of change requires both initial and final tests.
>
> 8. Scientific measurement is fair only when we measure amount of change produced in a standard time. This requirement can be satisfied.
>
> 9. Scientific measurement is fair only when we measure the amount of change in standard pupils, the Accomplishment Quotient is included in the plan . . . because this is a device for converting pupils no matter what their intelligence into standard pupils.
>
> 10. Scientific measurement is fair only when the measurement is complete. Absolute completeness would require a measurement of changes made in children's purposes as well as their abilities. Absolute completeness is, of course, impossible, and is in fact not necessary; partly because a chance sampling of the changes made will be thorough enough and partly because teachers' skill in making desirable changes in, say, reading is probably positively correlated with their skill in making desirable changes in, say, arithmetic (McCall 1922:150-152).

While there are many forces operating on the schools and many special interest groups concerned about educational effects, these must be secondary to any study in education. The first and foremost concern of any research project in education must answer the question: how will children be affected?

BEST COPY AVAILABLE

Children are what education is all about.

This study was concerned with the relationships of personality and motivational variables to music teaching success. Recognition was made of the fact that not only are personality behavioral dimensions multivariate in nature but so are the music teaching situations commonly found in various educational settings in the United States.

The Problem

The purpose of this study was to investigate the relationships between personality and motivational variables and music teaching success criteria. Answers to the following questions provided partial answers to the problem:

1. What are the relationships among teacher accountability criteria used to define music teaching success:

    a. class mean residual gain scores of pupils
    b. teacher ratings by pupils
    c. teacher ratings by principals
    d. teacher ratings by supervisors
    e. teacher ratings by peers
    f. teacher self ratings

2. What are the relationships and interrelationships of the personality variables measured by the Sixteen Personality Factor Questionnaire and the motivational variables measured by the Motivation Analysis Test to music teaching success?

3. Do personality profiles differ for successful and unsuccessful male and female music educators?

4. What are the implications of the use of personality measures for counseling and advising undergraduate music education students?

Delimitations

1. This study was limited to the psychological factors measured by the Sixteen Personality Factor Questionnaire (16PF) and the Motivation Analysis Test (MAT) by Raymond B. Cattell and others and published by the Institute for

BEST COPY AVAILABLE

Personality and Ability Testing (IPAT).

2. Only upper elementary and secondary school music teachers and senior undergraduate music education students were used in this study.

3. Pupil growth was limited to the music education behavioral objectives measured by selected subtests from the Music Achievement Test I-IV by Richard J. Colwell and published by Follett Educational Corporation.

4. Rating evaluations were limited to those behaviors measured by the Illinois Teacher Evaluation Questionnaire (ITEQ) by Richard Spencer and published by the Office of Instructional Services, University of Illinois.

Assumptions

The researcher made the following assumptions with regard to this investigation:

1. Participating music educators and music student teachers were teaching and student teaching at the level and speciality most conducive for achieving music teaching success.

2. Male and female music educators teaching in different sections of the United States had equal potential for achieving music teaching success.

3. Differences among schools and pupils due to organismic, experiential, and demographic variables could be statistically treated in order that participating music educators in widely varying music teaching situations could be rank ordered in music teaching success according to the criteria.

The Sample

This study was divided into two main samples: music educator and music student teacher. A total of 290 music educators participated of whom 209 were used in the final analyses. These music educators were located in 16 states and 80 school districts in the United States and Puerto Rico.

Music student teachers included 73 senior music education majors attending

7 teacher training institutions: University of Illinois, Eastern Illinois University, Western Illinois University, Illinois State University, Millikin University, Elmhurst College, and Oklahoma City University. All data were gathered from the music educator sample during the 1971-72 school year as well as for 72 of the music student teachers. The Oklahoma City University music student teachers comprise those students who student taught during the fall semester, 1972-73 school year.

The music educator sample was selected from music educators attending workshops and inservice classes at the aforementioned universities during the spring and summer terms of 1971. In addition, personal friends and acquaintances of the investigator throughout the United States and Puerto Rico were invited to participate in this study.

Music administrators in 43 large school districts in the United States who received the Bulletin of the Council for Research in Music Education were contacted requesting their cooperation in obtaining participating music educators who were paid a $15.00 gratuity. Meetings were held at each of the universities and in the large school districts inviting both the music educators and music student teachers to participate in this study. Because of the nature of personality information, each participant had an opportunity to decline to participate after the initial meeting and to drop out throughout the data gathering stages. Follow-up meetings provided for the gathering of data, testing, and answering questions concerning the study.

The purpose of offering the monetary gratuity to a certain portion of the participating music educators was an attempt to obtain subjects teaching music in socially disadvantaged schools and to attract teachers with widely varying abilities, years of teaching experience, and ages. Three large school districts through the influence of their music administrator cooperated fully with this study, and all the music teachers in the districts participated.

In order to validate the significance of regression equations, it is mandatory to use the equations derived from one sample of subjects on another like sample of subjects. Therefore, 20 percent of the music educators served for cross validating the regression equations. Furthermore, the main purpose of the music student teacher sample was for cross validation. If the equations could be cross validated on prospective teachers, the significance of advising and counseling undergraduate music education students would be accentuated. The main focus of this study was on the music educators because before one can predict potential music teaching success, he must first define music teaching success and investigate the relationships of the personality and motivational variables to the music teaching success criteria.

Definition of Terms

Average Music Educator is a person who is in the middle 50 percent of the sample according to the success criteria.

College Supervisor is the college or university representative who supervises the music student teacher.

Cooperating Teacher is the school district supervisor of a student teacher.

Music Performer is a musician who receives his primary income by performing professionally.

Peer Teacher is a colleague of the music educator and may teach any grade level or subject matter speciality in the school.

Personality Dynamic Dimension refers to a person's interest, drives, feelings, attitudes, and sentiments which provide the impetus for behavior. These personality dimensions are somewhat changeable due to environmental and sociological factors and are inherently less stable than temperaments.

Personality Temperament Trait Domain includes the basic personality characteristics and source traits such as anxiety, paranoia, social adjustment, economic conservation, intelligence, and so forth. These traits will not, under normal circumstances, change drastically from college throughout life.

Pupil Gain refers to the amount of gain made by pupils or a music achievement test and is arrived at by subtracting the pretest score from the posttest score.

Rating Evaluations refer to subjective judgments of teachers made by persons who know them well.

Residual Gain refers to the amount of gain made by pupils on a music achievement test and is determined by subtracting the raw gain score from the predicted gain score. This is a deviation or residual gain score.

Successful Music Educator is a person who is in the upper 25 percent of the sample according to the success criteria.

Unsuccessful Music Educator is a person who is in the lower 25 percent of the sample according to the success criteria.

Need For The Study

Many studies have investigated variables relating to potential success in college to some degree of usefulness. The ability to matriculate, however, does not guarantee that the graduating senior will become an effective teacher in the schools. What better predictor is there than to use previous achievement as an indication of potential achievement and success? One of the best predictors of potential academic success in college is the high school centile rank; yet there is no method currently available for ranking preteaching success in order to predict potential teaching success. For this reason, success in college as indicated by the grade point average is not very useful for estimating potential teaching success in the schools.

Schneider and Cady discuss the problem of selection in music teaching:

> Existing selection practices for prospective music teachers are not valid. Most are concerned primarily with the musicality of the student . . . No measure has been developed which will account for the non-musical variables which confront the teacher in the classroom and for which he must be educated (Schneider and Cady 1965:281).

One cannot envision a single measure or evaluative device which will effectively predict potential classroom success. Personality and motivational factors may account for a great deal of teacher effectiveness variance (Cattell 1965:314). Nonachievement variables appear to affect teacher decisions to remain in the profession or to quit. "The major causes for music teacher failure, however, are personality weakness, emotional instability, and tactlessness" (Schneider and Cady 1965:280).

Brown surveyed school administrators and music teachers finding the following music teacher weaknesses (in rank order): discipline, teaching skill, leadership, cooperation, emotional stability, tactfulness, academic knowledge, piano skill, conducting, enthusiasm, disposition, performing ability, basic music knowledge, music sensitivity, personal appearance, culture, morals, and intelligence (Brown 1955). The emphasis was on personality and motivational factors: leadership, cooperation, emotional stability, tactfulness, enthusiasm, disposition, and personal appearance. "Perhaps the most important and promising result of this survey is the conclusion that music teachers have no prevailing weaknesses different from those of teachers in other fields" (Brown 1955).

Ehlert found that lack of musical training and musical ability were not important causes of music teacher failure. Assessing the opinions of music supervisors and school district superintendents, he found the following 6 causes for the failure of music teachers:

1. Lack of personality
2. Lack of teaching ability
3. Lack of disciplinary ability
4. Lack of personal control
5. Lack of cooperation
6. Lack of interest in teaching (Ehlert 1951:36-38).

While "lack of personality" may be somewhat elusive in meaning, these 6 causes for failure relate directly to personality and motivational dimensions. The comparisons between the two surveys are interesting. Both found personality factors to be crucial to music teaching success in the opinions of those in charge: music supervisors, music teachers, and school administrators. "Cooperation" and "interest" are factors reflecting a person's personality structure.

A. S. Barr who dedicated his life to research defining successful teacher traits arrived at the following list of 17 types of causes for failure among school teachers:

1. Lack of control over the technique of teaching
2. Lack of ability to maintain order and discipline
3. Lack of master of subject matter
4. Lack of intelligence
5. Lack of effort
6. Lack of initiative
7. Lack of adaptability
8. Lack of common sense
9. Lack of physical ability
10. Lack of standards of teaching efficiency
11. Lack of ability to carry on
12. Lack of singleness of purpose
13. Lack of sympathetic understanding of pupils
14. Lack of social background
15. Lack of knowledge of what pupils can do
16. Lack of personality
17. Lack of moral standards (Barr 1938:359-361).

After 36 years of research the lists of why teachers fail are similar and appear to be centered around personality and motivational variables. While music teacher weaknesses may be no different from those of teachers in other fields, one may find differences in personality and motivation within subgroups of teacher specialities.

A publication of the National Council of State Supervisors of Music discusses the qualities and competencies needed for successful music teaching and states that music educators must possess something more than just performance and musical skills:

> The development of music teacher competencies should result from the total program of the teacher training institution. The demonstration of these competencies, rather than the passing of a course, should be the deciding factor in certification. This means that proficiency tests, practical applications of historical, theoretical, and stylistic techniques, and advanced standing procedures should be employed, that screening procedures for admission to the program be enforced, and that an adequate means of final assessment should be developed and implemented (National Council of State Supervisors of Music 1972:9).

The supervisors recognize that not only are music competencies required but that personal and professional competencies are apposite to engendering music educators better able to meet the diverse needs of today's teaching situations. While most of these suggestions have been based upon empirical evidence, these personal qualities are of interest for research investigations.

> Music educators must (1) inspire others, (2) continue to learn in their own and in other fields . . . (3) relate to individuals and to society . . . (4) relate to other disciplines and arts . . . (5) identify, and evaluate new ideas . . . (6) use their imagination . . . (7) understand the role of a teacher . . . For example, the ego-satisfaction of the music student in college is gained through performance, while that of a music educator will be gained largely from creating opportunities for students' musical expression (MENC 1972:48).

Emphasized are the personal qualities of leadership, attitudes of intellectual curiosity, empathetic feelings, positive attitudes toward music and the related arts, scientific inquiry which includes the ability to utilize technological, experimental, and exploratory developments in musical composition, teaching procedures and aids, and sound generating devices; creative ability with musical materials and learning problems, and creativity in relationships with colleagues (MENC 1972:48). While one may wonder where to seek such an irrefragable soul, it is important to acknowledge that the Music Educators National Conference in recent publications has emphasized the need for going beyond mere musical preparation of potential teachers and to look at the total music teaching situation as well as the qualities needed in order to survive successfully in that environment. The ability to predict potential music teaching success becomes crucial.

More teachers are being graduated by institutions of higher learning than can be placed in teaching positions in the schools under current hiring practices. In Illinois for the 1971-72 school year, there were approximately 5,000 more qualified graduates than vacant teaching positions. Charters states the case most succinctly for the University of Illinois' graduates in the teacher training program in all subject matter areas:

> Computations based on University of Illinois graduates over a ten year period suggest that 40 percent of those qualified to teach never take public school jobs. One half of those entering the teaching profession drop out after two years of teaching. Only 150 of every 1,000 graduates qualified to teach are still teaching after five years of experience. Only 96 of the original 1,000 graduates are teaching during the tenth year in the teaching profession (Charters 1956:253).

It is not suggested that most of these teachers drop out for reasons of failure. It may be better for all concerned to have some assessments at entrance to college or during the teacher training program which may show potential for teaching success.

In music education, the figures are better. At the elementary level, 66 percent of the 933 elementary music education students completing their training between September 1, 1970, and August 1, 1971, were employed in their field of study by November, 1971. This marks the highest percentages in any elementary school field. At the secondary school level, 56.2 percent of the music education majors entered the teaching profession. In comparison to earlier years, the situation for employment of elementary and secondary music education graduates is bleak. During the peak year of 1958, 74.5 percent of the music education graduates entered classrooms immediately after attaining their degrees but this percentage figure has declined steadily. The reasons for this are not clear.

The NEA survey indicates that, taking into consideration both graduates who are entering teaching for the first time and teachers returning to active status after a period of absence, an adequate supply of teachers in music, art, English language arts, foreign languages, home economics, social studies, and regular elementary education will be available. These data are based on present staffing practices and do not reflect the demand projected if schools were to meet a minimum standard of quality, reduce overcrowded classes, replace teachers with substandard qualifications, and offer special instructional services. This research reveals over 1,800 cases where elementary level programs in music, art, foreign language, home economics, remedial reading, physical and health education, industrial arts, and science were discontinued for financial reasons; and, at the secondary level, more than 1,300 school systems eliminated their offerings in agriculture, art, driver education, foreign language, home economics, industrial arts, music, reading, and/or vocational office training (Music Educators Journal, May 1973:68-70).

The personality characteristics of men and women music educators are unknown. Previous investigations have correlated personality variables to music teaching success criteria and have found some factors significantly related. But, just because significant correlations or significant t-tests on personality variables between groups or among criteria are found, one cannot assume that these variables are useful in estimating potential music teaching success. Personality behavioral dimensions are multivariate in nature as are the criteria of music teaching success and the music teaching situations commonly found in various educational settings in the United States. These areas must be considered within a research design intended to investigate the relationships of personality and motivational variables to music teaching success criteria. No study has adequately defined the personality dimensions, relationships and interrelationships, useful for estimating potential music teaching success for counseling and advising prospective music educators.

David G. Ryans says:

> The identification of qualified and able teaching personnel, therefore, constitutes one of the most important of all educational concerns. Obtaining capable teachers is an intrinsic interest and obligation of education (Ryans 1960:1).

Howard Mitzel agrees: "The task of identifying effective teachers (or effective teaching) is crucial to teacher education, certification, selection, and promotion . . ." (Mitzel 1960:148). With this last point, the significance of this study rests.

CRITERIA OF MUSIC TEACHING SUCCESS

Before research can be instigated to investigate the relationships of personality and motivational variables to music teaching success criteria, definitions must first be offered. The definitions or criteria of music teaching success must be capable of replication and apply to a wide variety of music teaching situations. Therefore, the criteria of music teaching success were under investigation in the study.

Criterion validity was a major concern. Psychometricians recognize that value judgments ultimately have to influence the criteria. "Is the good teacher the one who is most loved, or the one who teaches the most painfully necessary lessons" (Cattell 1970:238)? It is no easy solution to include every aspect of the criterion domain.

> It is reasonable to expect that the optimum personality profile will vary somewhat with the nature of the teaching, e.g. elementary school, high school, college . . . The most desirable criterion - subject and personality progress in the students - still eludes an adequate scale research (Cattell 1968:168).

Brophy offers evidence tnat certain identifiable teacher behaviors produce varying results with different kinds of pupils. Teachers who were successful with high SES pupils were demanding and dogmatic while teachers who were successful with low SES pupils were more patient and flexible (Brophy 1973:250-251).

This study sampled the domain of music teaching success. Six criteria were selected for investigation because they appeared to be the major concerns in evaluating music teaching success:

1. Residual gain scores of pupils
2. Teacher ratings by pupils
3. Teacher ratings by building principals
4. Teacher ratings by music supervisors
5. Teacher ratings by peers
6. Teacher self ratings

These six criteria, while not wholly inclusive, were representative of major areas in music teacher evaluation.

> . . . Music directors and supervisors will have to assume a more aggressive role in identifying goals that can be measured behaviorally so that specific outcomes in music may be demonstrated and accounted for . . . They can no longer rely on euphemisms, platitudes, and cliches to describe the inherent value of their programs (Klotman 1973:83).

Residual Gain Scores of Pupils (1)

One of the major purposes of music education is to affect learning. If learning occurs, this achievement can be objectively measured to some extent.

Barr reports that the first step in the analysis of a complex learning and teaching situation is the evaluation of pupil growth and achievement in terms of the purposes of education (Barr 1938:357). Contrary to the opinions of Lucio and McNeil, pupil product measurements for teacher accountability criteria have received attention and systematic study (Lucio and McNeil 1969:240). Much effort by Barr and others in the 1920's and 1930's was expended in developing measures and techniques for evaluating pupils' progress as aspects of teacher success. Recent research efforts by the U. S. Office of Education, the National Institute of Education, National Assessment, and programs such as contract teaching have renewed inquiries into pupil measurement.

Tests for the pupil gain criterion were compiled from the standardized Music Achievement Tests I-IV by Richard J. Colwell. The Music Achievement Tests include subtests on pitch, interval, major-minor, and meter discrimination, auditory-visual discrimination, feeling for tonal center, tonal memory, melody, pitch, and instrument recognition, and identification of style, texture, and chords. The reliability is high, and the validity has been established by correlations with other tests, by performance ratings of students made one year after taking the test, and by a variety of teacher ratings of students. The norms were developed from a carefully, stratified sample of more than 20,000 pupils representing 158 school districts in 43 states and according to Hoffer, "a far greater number than used in standardizing any other music test" (Hoffer 1973:114). "The most notable published tests of this type [music achievement] are the Music Achievement Test(s) by Richard Colwell and the Achievement Tests in Music by William Knuth" (Hoffer 1973:113). Furthermore, much additional research is currently under way in the area of predictive validity of the MAT Battery.

Whybrew considers the Music Achievement Tests to be "an important contribution to achievement testing in music" and concluded by saying that:

> "perhaps the outstanding feature of the Music Achievement Tests is that they emphasize the auditory aspect of music . . . Colwell has made a very interesting contribution to the literature of music tests, and it is hoped that research studies and use of the battery will provide more information about its advantageous utilization (Whybrew 1971:166,170).

Lehman concurs: "the . . . Music Achievement Tests represent a most promising attempt to provide an objective measure of the extent to which pupils are acquiring some of the basic skills and concepts music teachers presume they are teaching" (Lehman 1968:68).

Lehmar in reviewing the Music Achievement Tests for the Seventh Mental Measurements Yearbook states:

> The tests have been carefully constructed. They include several item types not previously used on standardized tests in music. The manuals are complete, the explanatory materials adequate, and the claims made for the tests are realistic.
>
> The Music Achievement Tests are useful in providing students and teachers with evidence of progress toward meeting the objectives

> of the curriculum and in providing information of use in student guidance and in program evaluations.
>
> Despite its limitation in assessing individuals, MAT is the best, most comprehensive, and most widely useful standardized achievement test battery in music currently available. Ultimately its usefulness to the teacher will be determined by the degree of correspondence between the skills that he is attempting to teach and the skills measured by the tests . . . MAT is an important and valuable general purpose test for use in a wide variety of situations (Lehman 1972:247-249).

In order to assure music achievement test validity for music classroom use, each participating music educator listed 5 music behavioral objectives for which he taught. Each participant selected his 5 objectives from a list of 13. From the teachers' 5 objectives, the investigator selected 3 which he felt would be most appropriate for producing gain. Teachers of middle and secondary school pupils were provided the 3 most difficult test sections of the 5 which they selected. Had they been given easy test sections, e.g. those more appropriate for elementary pupils, their middle and secondary students would have had a tendency to score high on the pretest allowing little room for growth. A second consideration was to provide the music educators at similar levels - elementary, middle or secondary - with test sections that were selected by the majority of teachers at the 3 levels so as to reduce the total number of specially prepared music achievement tests.

The Follett Educational Corporation graciously consented to this special use of the MAT tests and to the duplication and construction of answer sheets. No more than 30 minutes of classroom testing time was required for each pre and posttest session.

The reliability of the MAT series was computed by split half and KR 21. Reliabilities for test sections range from .69 to .94 and are therefore well suited for use in this investigation.

Today, there is a controversy about the validity of gain, change, or difference scores when used for testing hypotheses about treatment effects in experimenta[redacted] or quasiexperimental design. The common ap[redacted]ches include raw gain, analysis of covariance (ANCOVA), residual gain which includes partial and part correlations and multiple regression techniques, and other correlational procedures using gain as a final status criterion.

The gain criterion selected for this study was one suggested by O'Connor which included certain organismic, demographic, and experiential variables within the residual gain equation (O'Connor 1972:89). In this technique, the class is the unit of measurement rather than one individual pupil. Brophy has also used this technique in his Texas studies of elementary teacher effectiveness (Brophy 1973:151).

Because the groups used for computing the residuals were not randomly assigned in the study but were rather intact music classes, a multiple regression equation was formulated which used the following variables to compute the class residual gain scores:

X_1 = (variable created from variables X_2 and X_4 through X_6)
X_2 = (a variable for the number of class meeting times per week/3)
X_3 = (the pretest mean, initial status mean of the class)
X_4 = (a variable for the number of minutes per session for each music class/40)
X_5 = (a dummy variable for intervals between pre and posttesting/20)
X_6 = (a variable for school classification: socially advantaged, socially disadvantaged, or "in between"/2)
Y = (the posttest mean of the class)
Q = (the residual)
b = (the unstandardized regression coefficient or b weight)

Variable X_1 was created by the following equation:

$$X_1 = \frac{X\ X_2\ X_4\ X_5\ X_6}{150}$$

The 150 denominator was used to factorially reduce the size of the created variable in the equation and represented an approximate mean score for the numerator variables. Mathematically, any constant could have been used as a divider without affecting the nature of the variable in the equation.

BEST COPY AVAILABLE

Variable X_2 (meeting times per week) was considered to be of sufficient importance to be entered into the equation as a separate covariate.

The following equations were used to derive the class mean residual gain score for each participating music educator:

$$\hat{y} = b_o + b_1N_1 + b_2N_2 = b_3N_3$$

$\hat{y} - \bar{y} = Q$, the residual or deviation class mean gain score

Because each participating music educator selected 5 objectives of which 3 were evaluated by the specially prepared music achievement tests, one might question the validity of rank ordering teachers from high to low according to residual gain scores when different tests were used. Since the residual gain score of the class was a theoretical construct of music teacher accountability, it appeared best to use test sections from the Music Achievement Tests which pertained to the stated objectives of each music educator. Whether this test was too easy or difficult, appropriate or inappropriate for the pupils was of concern to this investigation only as it pertained to the teacher himself. If the teacher erred in selecting the music class for the music achievement testing or in stating his objectives which affected the music achievement test sections used with his pupils, this was an indication of the teacher's ability to produce or not produce gain as measured in this study. The assumption was that the music educator knew his situation best. Gain in this sense is not measuring the pupils' achievement on certain musical behaviors per se but instead serves as an indicator, a theoretical construct, of a music teacher's ability to produce musical learning - pupil achievement - as measured by the MAT test sections.

The alternative approach would have been to use one test for all the teachers; but since musical objectives differ, test validity would suffer. Another approach would have been to stratify the pupils within grade and music class categories throughout the sample and compute individual pupil residual gain scores. If this were done, pupils' scores would have been standardized across the country for each subtest or test section thus norming the pupil sample. This would have been unfair. Class residual gain scores were computed using the aforementioned equations which accounted for some demographic, organismic, and experiential variables.

Because there were 4 periods or times of pretesting - 2 in the first semester and 2 in the second semester - it was necessary to check for differences in residual gain that could be attributed to pretesting time differences. This could not be assumed to be a linear relationship and thus treated in the residual equation. Through ANCOVA and t-test procedures, differences within the 4 pretesting groups were investigated. Considering the teacher response variable as far as pre and posttesting times and intervals over which this project had little control, this technique seemed to be a fair approach to the teacher accountability criterion of residual gain.

The residual gain technique is questionable, but at the present time, it seemed to be the best technique available. Granted, the gain construct criterion was very limited when one considered the total amount of objectives and behaviors learned in any given music teaching situation. In any event, the residual gain technique was used in order to investigate objective music achievement test uses with music classes and relationships to subjective rating criteria as well as personality and motivational variables.

The residual gain procedure accounted for the possibility that the various music achievement test sections differed in ease of teaching or in difficulty of measurement. By fitting a straight regression line for all the pupil scores within the class, these factors were somewhat accounted for in the design. The residual gain technique best accounted for the factors of test fallibility and regression effects.

Teachers were mailed testing packages including specially prepared and duplicated answer sheets and a test tape including three MAT subtests. These packages including sheets and tapes were returned to the investigator after pretesting and were remailed and returned for posttesting. Follow-up letters to the music educators prodded them to pre or posttest after two and four week delays. The music achievement tests were scored by hand by the investigator and his staff. Teachers were rank ordered within the above mentioned subgroups.

The investigator accepted the views of McNeil and Popham:

> A focus on pupils reveals far more about the effectiveness of teachers than does direct study of teachers themselves . . . Support for the position that the ultimate criterion of a teacher's competence is his impact upon the learner has been offered by a number of individual researchers as well as professional associations . . . (McNeil and Popham 1973:218).

Ratings of Music Teaching Success

Five ratings of music teaching success were used in this investigation. The ratings were the opinions and attitudes which the rater had about the teacher being rated. The Illinois Teacher Evaluation Questionnaire (ITEQ) is designed to assess the perceived attitudes of pupils about their teachers. This form has a total of 40 items and was amended to a 23 item form in order to obtain ratings from building principals, music supervisors, peer teachers, and music educator self ratings.

ITEQ is published by the Measurement and Research Division of the Office of Instructional Services, University of Illinois. Pupils' opinions are assessed in four subscore areas: teacher, teacher's competence, interest to the class, and teacher's attitude. While the subscores were of interest to participating music educators and music student teachers, only the total scores were used for the pupil rating criterion.

The first subscore (teacher) refers to general impression of the teacher. The second (teacher's competence) gives an indication of perceived evaluations of teacher's knowledge and general overall competence. The third (interest to the class) gives an evaluation of the time spent in class and doing outside work for the class. The fourth subscore (teacher's attitude) is an indication of a person's perceived evaluations of the teacher's attitude and philosophy toward teaching.

ITEQ is a forced choice instrument requiring the respondent to strongly agree, agree, disagree, or strongly disagree with each of the 40 statements. Scores ranged from 1 to 4 and were standardized within each rating category to a range from low to high of 1 to 10 with a mean of 5.50 and a standard deviation of 2.00.

ITEQ has been found to be a reliable and valid instrument. Reliability or consistency coefficients range from .59 to .70 for subsections with the total score of .62. Individual class reliabilities for participants in this study for total test score were all in the .90's. The intercorrelations range from .72 to .99, and, while the instrument does not have the high reliability as other objective tests, it was deemed adequate for use in this study.

ITEQ was developed from the Illinois Course Evaluation Questionnaire (CEQ) which was designed for use with university level courses. ITEQ is designed for use in the secondary schools and at the time of this project was being investigated extensively. CEQ's reliability using the split half methods when corrected for length by the Spearman Brown formula yielded reliabilities of .92 and .93 respectively in two studies. It can be expected that ITEQ's reliability is equally consistent (Aleamoni, Memorandum 127:5).

While there have been many proposals in the past few years to evaluate instruction, most contain similar elements or areas of concern such as pupil, peer, and supervisor ratings. If, however, one looks for actual working models of instructional evaluation, it is apparent that schemes involving systematic ratings by peer, supervisor, and self, as well as of material, content, and so forth are very seldom actualized. More often than not, the student ratings of instructor and instruction at the university level appear as the only elements in any working model (Aleamoni 1972:1).

A USOE funded research study investigated the relationship between student gains and ratings by trained impartial observers. Heil and others studied gains in relation to (1) the students' intelligence scores, (2) the students' emotional patterns, (3) the teachers' personality patterns, (4) the teachers' knowledge in the liberal arts and education, and (5) the students' socioeconomic status (Heil and others 1960:10).

Through ANCOVA techniques, they found the following: (1) For all children and for most academic subjects, the self controlled teacher was the most effective while the fearful teacher was the least effective. (2) The self controlled teacher produced greater gains in her students. (3) There were no significant differences in gains between children of different social classes. (4) The professional education course grades of the teacher as well as the ratings by the observers had no relationship to the teachers' effectiveness as indicated by the gains criteria. (5) There was a negative relationship at the .05 level between the teachers' tested knowledge in liberal arts and their effectiveness as measured by the gains criteria (Heil and others 1960:50-56). Heil concluded, "That not only is teacher knowledge an unimportant variable as a criterion of teacher effectiveness but also observers' ratings, per se, are next to worthless in this respect" (Heil and others 1960:77).

Hellfritzsch found no significant correlation between the measured student gains and the supervisory ratings of the teachers' effectiveness (Hellfritzsch 1945:199). Morsh in an Air Force study concluded, "High correlation found between supervisor and fellow instructor rankings, coupled with the fact that neither of these measures correlated highly with the student gains criteria suggests that both supervisors and peers judged instructors on the basis of factors other than teaching effectiveness as measured by the gains criteria" (Morsh 1956:45). Possibly what was operating were simply two domains of success - objective and subjective.

The literature from the 1920's to the present supports the thesis that there are at least two domains of teaching success: objective in terms of gain scores and subjective in terms of rating evaluations. The Barr studies support this as well as those cited above. This project, investigated these two domains, the relationships between them, and the relationships and interrelationships to personality variables.

Teacher Ratings by Pupils (2)

ITEQ was used to assess pupil attitudes. All participating music educators administered these ratings in May, 1972. The subjects were encouraged to have their building principal, counselor, another teacher, or a student actually administer the ratings to their pupils. Music student teachers' pupil evaluations were gathered after the student teacher had completed the student teaching experience and were administered by the cooperating teacher.

Few studies have investigated pupil judgments of teachers as measures or criteria of teaching effectiveness. Albert concluded that pupil judgments are consistent, reliable, valid, practical, and inexpensive [unless one used a standardized instrument]. He found little agreement between administrator judgments and pupil judgments (Albert 1941:274).

Silberman reports an item found by researchers at a graduate school of education. High school students rated student teachers. These evaluations were compared with those made by the college supervisors. "To their surprise, the researchers found that students' evaluations were more consistent than those of the supervisors" (Silberman 1970:454). While this indicates, perhaps, greater reliability for the students' ratings, Silberman did not report the relationship between the supervisor and student evaluations.

Most of the research comparing pupil or student ratings of instruction to other ratings has been conducted at the university level and with the CEQ although other standardized instruments are available, e.g. *Purdue Rating Scale* and so forth. Stallings and Spencer (1967), using 10 judges consisting of measurement specialists and teaching assistants attempted to determine the relationship between judges' ratings and student ratings. The judges' ratings were based solely on video clips which were rated on a three point scale. Student ratings were based on the total CEQ scores at the end of one semester. A significant Spearman rank order correlation ($p=.70$) was found between the CEQ ranks and the average rating ranks. Swanson and Sisson (1971) obtained the same Spearman rank order correlation ($p=.70$) between peer ratings and student ratings using the CEQ total score (Aleamoni 1972:2).

The relationship between CEQ ratings and instructor's rank, class size, and level of course was examined by Graham (1972) using MANOVA techniques. As hypothesized, there were no significant differences in ratings assigned by students in small (1-20), medium (21-40), and large (over 40) classes ($p<.13$). In addition, significant size by level (freshman, sophomore, junior, senior, graduate) ($p<.02$) interaction effects were found. Discriminant functions computed for effects found to be significant yielded information concerning the extent and direction of these significant differences.

The contention has been made that students rate classes and instructors differently according to the time of the class meetings. Differences among classes meeting at various times were found to be statistically significant but of little practical value since the statistical significances were probably due to the large sample sizes (Aleamoni 1972:3).

Colleague and student ratings were not found to be significantly related to the instructor's research productivity: however, the colleague ratings were significantly related to academic rank ($r=.20$) indicating that the reputation of the instructor may have an influence on colleague ratings. The UCLA Task Force on Evaluation of Teaching emphasized that "a basic element of evaluation should include a regular survey of teaching effectiveness as perceived by students" (Aleamoni 1972:4).

Centra, in a study of student evaluations of college instructors, used the "Student Instructional Report," a 23 item form that elicited students' judgments of such aspects as the organization of the course, the pace of the course, the instructor's helpfulness or availability to students and the clarity of objectives and presentations. He found differences in pacing among the courses with humanities teachers less likely to inform students of how they were to be evaluated and less agreement between announced objectives and what was actually taught in these classes. Students rated female teachers more favorably than male teachers on several items, particularly those dealing with teacher-student interactions. A comparison of students' ratings of instruction with teachers' self reported ratings disclosed only a modest positive relationship between the two sets of evaluations. This suggests that student ratings do provide instructors with information they do not already have about their teaching. However, the major finding was that student feedback did produce some changes in instruction when the student ratings were significantly lower than the teachers' self ratings. The greater the discrepancy, the greater the likelihood of instructional change. He also found that student ratings of instruction are positively skewed, well over half of a group of typical instructors scored above the midpoint on the scale (Centra 1973:12-13).

Building Principal Ratings (3)

Building principal ratings of teaching effectiveness are a common criterion of teaching success. Most research with personality variables and teaching effectiveness has used some form of principal or superintendent rating as one of the success criteria.

Borkowski found a general agreement to the validity of administrator ratings (Borkowski 1967:23-25). Barr in his studies found relatively high correlations between principal, superintendent, and supervisor ratings.

Building principals rated the participating music educators in May, 1972, with some follow-up in summer and fall. Ratings were mailed directly to the building principals, anonymity was assured, and the principals classified their schools as socially advantaged, socially disadvantaged, or "in between."

Music Supervisor Ratings (4)

Ratings were obtained from the music supervisors of participating music educators in May, 1972, with follow-up in the summer. These were mailed directly to the supervisors and anonymity was assured. The supervisors also classified the schools in which their music teachers were situated as socially advantaged, socially disadvantaged, or "in between." When the supervisors' classification disagreed with the building principals' classification, the "in-between" category was used for classification. Also, while this was not a choice on the form, many of the administrators indicated that their schools were 50 percent of each.

Music supervisor ratings are most common in music education research. While there has been a shift in the role and power of the music supervisor today, he still maintains an influential position within the music staff. More than 75 percent of the participating music educators had music supervisors.

Peer Teacher Ratings (5)

Fellow teachers may be able to effectively evaluate the teachers with whom they associate even though they may never actually see the music teacher teach music. It seemed appropriate to investigate this criterion and to determine the relationships to the other criterion variables.

In May, 1972, each of the participating music educators received a packet of 5 ratings. These ratings along with a stamped, addressed envelope were given by the music educators to 5 teachers who they considered qualified to evaluate their music teaching success. These were returned directly to the investigator by the peer teachers.

Teacher Self Ratings (6)

Self ratings may be a valid means for identifying effective music educators. Few research investigations have studied this criterion at the secondary school level but the research at the college level, as previously discussed, indicated usefulness when discrepancies occur. The self ratings were completed in May, 1972, with some follow-up during the summer.

Music Student Teacher Ratings

The following ratings were gathered for student teaching success criteria upon conclusion of the student teaching experience: pupil, cooperating teacher, college supervisor, peer student teacher, and self. Because this group was to serve as a cross validation sample, it was hoped that some comparison could be made between the music educator ratings and the music student teacher ratings.

Statistical Procedures

Standard scores were produced for each rating to transform the scale from 1 to 4 (low to high) to 1 to 10 with a mean of 5.50 and a standard deviation of 2.00. Because not all music educators had music supervisors and missing data were anticipated, missing data rating scores were predicted. A step wise multiple regression approach using all the rating scores was adopted which predicted up to two missing rating scores for teachers with incomplete criterion data. Those participating music educators and music student teachers with more than two missing ratings were eliminated from the final analyses.

Correlational analyses were undertaken to investigate the relationships of the ratings. A principal components factor analysis without rotation weighted the ratings for the music educator and music student teacher samples. The weights from the first principal component were used in both cases and the samples were rank ordered from high to low. For some preliminary analyses, the subjects were stratified into successful, average, and unsuccessful groups. With this procedure, a weighted single rating score was derived for each participating music educator and music student teacher.

Analysis of variance investigated whether the standardized rating scores for each of the 5 rating categories differed among teaching and speciality levels: elementary school instrumental and vocal music, middle-junior high school instrumental and vocal music, and high school instrumental and vocal music. There was no indication in the literature to suggest that there were any relationships between music teaching success and the specific music teaching speciality. Also, correlational analyses were undertaken between the residual gain scores criterion and the rating criterion to investigate significant relationships.

The male and female music educator samples were rank ordered and for some preliminary analyses divided into thirds -- successful, average, unsuccessful -- according to the following criteria: Gain (G), Rating (R), Gain plus Rating (G+R), and Gain minus Rating (G-R), a discrepancy criterion. This procedure provided for a number of ways with which to investigate the relationships within the criterion variables and to the personality variables. The basic matrix for groups was:

	High Gain	Average Gain	Low Gain
High Rating	1	2	5
Average Rating	3	4	7
Low Rating	6	8	9

PERSONALITY AND MOTIVATIONAL VARIABLES

Music Teaching Specialities

Bodenstab in a study of personality patterns and heterogeneity among teaching interest areas found psychometric differences among the grade level groups of elementary, special, and secondary; and subgroups of eleven teaching areas of which music was one. He said, "The greater than chance assignment to groups (three grade levels) was 17 percent and the greater than chance assignment to subgroups (eleven teaching interest areas) was 33 percent." He employed a MANOVA design which evaluated 12 variables of a psychometric instrument to the teaching grade level and interest areas. He concluded that the teaching profession is psychologically heterogeneous with respect to interest areas (Bodenstab 1966:319).

Music educators are a subgroup of the total population of educators. Wink investigated personality variables' relationships to achievement in music student teaching. Although he did not divide his sample according to sex, he indicated that music student teachers differed significantly from the national norms on the instruments he used (Wink 1970:234-241). Investigations of personality factors by Duda, Fosse, Funk, Morgan, Picerno, Turrentine, and Williams, to mention but a few indicated that the music teaching speciality (instrumental and choral) did not produce significant differences as far as the subjects' basic personality behavioral traits and patterns as measured by the Minnesota Multiphasic Personality Inventory, the Guilford-Zimmerman Temperament Survey, and the Sixteen Personality Factor Questionnaire.

Psychometricians agree that subgroups of teachers, e.g. music teachers, English teachers, history teachers, and so forth, should be divided by sex when personality is studied and that sex should be treated as a moderating variable or that separate groups should be stratified by sex.

> As we move up the educational ladder, we find no noticeable differences between junior and senior high school teachers, but a decided one between men and women, consistent in nature at both teaching levels . . . These differences may be due to simply the usual sex differences known to exist on these source traits (Cattell 1968:183).

There are significant differences between responses of males and females on personality and motivation tests. These differences have been treated in this investigation.

From empirical evidence, there was no reason to assume that teaching experience is a factor in music teaching success. Some very early studies of general elementary classroom teachers found differences between beginning and experienced teachers. While these may have been due to training variables of less than 4 years of college, there is no reason to believe that first year music teachers do not have the capabilities for being as outstanding in affecting pupil learning as teachers with 5 to 10 years' experience (Betts 1935:202).

Picerno found no significant relationship existed between judges' ratings of music teaching success and years of teaching experience (Picerno 1956: 250). Furthermore, Lutz in his personality study found that successful instrumental music teachers were younger than the unsuccessful ones (Lutz 1963:143).

Because music educators are often trained for teaching at all levels of public school music, the level was not a factor under investigation except with respect to the criterion variables. However, music teaching speciality was investigated with respect to the personality and motivational variables.

The 16PF

The 16PF is a unique personality questionnaire designed for many psychological uses. The test was first published by the Institute for Personality and Ability Testing (IPAT) in 1949. "Perhaps it befits the complexity of an instrument which has to measure the human mentality that it should have had a definite infancy, between the first and the 1956-57 edition; a childhood, between that and the 1961-62 edition; and an adolescence terminating in the present Handbook, with its associated new editions" (Cattell and others 1970:xix).

The 16PF has three important properties. First is the unusual comprehensiveness of coverage of the personality dimensions. Second is the orientation of the scales to functional measurement. This means that the scales were not formulated according to subjective or a priori concepts but are directly related to previously located natural personality structures related to the way personality actually develops. Third, these personality measurements become increasingly relatable to an organized and integrated body of practical and theoretical knowledge in the clinical, educational, industrial, and basic research fields (Cattell and others 1970:5).

The 16PF is based on a series of interlocking researches over 25 years, directed to locating unitery, independent, and pragmatically important source traits both in ratings and questionnaires. Cattell defines source traits as "factors (rotated to oblique simple structure) affecting large areas of the overt personality behavior . . ." (Cattell and others 1970:7). The source traits embodied in the test have been repeadedly replicated in basic personality research and related to systems proposed by others, i.e. Guilford-Zimmerman scales, Eysenck's scales, and the MMPI (Minnesota Multiphasic Personality Inventory). The items have been continuously improved and selected for higher loadings on each factor, consistent with good suppressor action and a wide sampling of behavior consistent with broad personality factors (Cattell and others 1970:8-9).

The 16PF investigators have related the trait factors to observer ratings, to higher order structures, and to various instrument factors and motivational distortion from which separation is needed. Research is rapidly providing evidence of broad profiles for a wide range of clinical, prognostic, and performance uses. Cross cultural studies have provided evidence of basically similar personality source trait structures.

IPAT publishes five versions of the 16PF with a sixth in development: *Forms A and B* for normal adult use, *Forms C and D* for average adults having less vocabulary abilities, and *Form E* with *F* under development for low literate adults. Four other scales measuring personality dimensions are available from IPAT: *Forms A and B* of the *High School Personality Questionnaire* for ages 12 to 16 years, the *Early School Personality Questionnaire* for ages 6 to 8 years, and the *Pre-School Personality Questionnaire* for ages 4 to 6 years. These tests allow for comparisons of personality measurements of persons at different scale such as Scale A not only has the properties of A as directly examined but also whether it relates to the not-A in the universe just as A does. "As A's relations to the not-A in the universe approach those for A, A and A approach logical identity" (Cattell and others 1970:38-39). Table I shows the direct and indirect concept validity estimates.

Concrete validity refers to the effectiveness and usability of the test instrument. The validity is as high as can be demonstrated anywhere for source trait measures of this length. If an adequate validity is reached, one is assured that requisite reliability also exists. The reason for using both forms of the 16PF was to increase validity for this study.

Reliability coefficients based on test-retest procedures are provided in Table I. Cattell discusses the value of concept validity as an approach to do more than make actuarial prediction.

> The meaning of the factor includes such information as that its variance in a particular group might be expected to be less than in some other group; that it has more or less constitutional than environmental determination; that its typical age curve in our society is of such and such a kind; that it might respond to a traumatic situation in such and such a way, . . . surely if we are to be more than 'psychological accountants' all this extra information is of considerable value and enables us to apply psychological laws, so that in fact our predictions from a

TABLE 1.--16PF PRIMARY FACTORS

Factor	Low Sten Score: 1-3	High Sten Score: 8-10	No. of Items Forms A & B	Reliability Coefficients A	B	A+B	A+B	Validity Coefficients A	B	*A+B	Indirect**
A	Sizothymia Stiff	Affectothymia Participating	10	81	75	89	85	79	78	90	96
B	Low Intelligence Dull	High Intelligence Bright	13	58	54	65	63	35	44	72	95
C	Low Ego Strength Easily Upset	High Ego Strength Calm	13	78	74	87	75	70	66	82	95
E	Submissiveness Accommodating	Dominance Stubborn	13	80	80	88	85	63	64	90	91
F	Desurgency Serious	Surgency Enthusiastic	13	79	81	90	78	83	79	93	96
G	Weak Superego Expedient	Strong Superego Conscientious	10	81	77	88	84	67	69	89	94
H	Threctia Shy	Parmia Venturesome	13	83	89	93	88	92	87	92	95
I	Harria Realistic	Premsia Overprotected	10	77	79	89	87	70	75	88	96
L	Alaxia Trusting	Protension Suspicious	10	75	77	87	76	49	63	85	91
M	Praxernia Practical	Autia Imaginative	13	70	70	82	71	44	73	85	74
N	Artlessness Forthright	Shrewdness Astute	10	61	60	76	74	41	60	84	83
O	Untroubled Adequacy Self Assured	Guilt Proneness Apprehensive	13	79	81	89	77	71	81	94	84
Q1	Conservative Conservative	Radicalism Experimenting	10	73	70	83	83	62	51	84	83
Q2	Group Adherence Group Dependent	Self Sufficiency Self Sufficient	10	73	75	85	81	70	70	88	90
Q3	Low Self Sentiment Self Conflict	High Self Sentiment Controlled	10	62	62	78	70	68	69	90	93
Q4	Low Ergic Tension Relaxed	High Ergic Tension Tense	13	81	87	91	78	57	59	87	93

1 - Reliability for Forms A and B indicated separately were test-retest, 4 to 7 day interval.
2 - Reliability for the combined Forms A and B were test-retest with a 2 month interval.
3 - Validity reported for each Form (A and B) are Direct Validity Coefficients.
* - Validity for A + B combined with computer synthesis scoring.
** - Indirect Validity Coefficients.

> test today to behavior a year hence are decidedly better than could be made on purely statistical grounds (Cattell and others 1970:263).

Criterion validity has been established for various occupations, for many industrial and clinical uses, and through numerous research projects with widely differing persons.

The 16PF is a good test which has been developed and revised in 25 years of research. Getzels and Jackson state:

> Too little has been done with the 16PF Test, with teachers as subjects, to accept without question the results that have appeared and to make possible a comprehensive evaluation of the method. In some cases, the findings of the several studies contradict one another. Nevertheless, the instrument has at least two specific advantages (aside from purely technical considerations). First, by providing scores on factors that are not purely evaluative (i.e. psychologically 'good' or 'bad'), the test encourages the use of hypotheses that are more sophisticated than those linking 'adjustment-maladjustment' or some dichotomous variable to the complex phenomena of teaching and of teaching effectiveness.
>
> Second, the instrument derives from an extensive program of both theoretical and empirical work carried out by Cattell and his associates over a number of years . . . The resulting body of concepts and findings would seem of considerable heuristic value for investigators intending to use the 16PF Test for studies of teacher personality (Getzels and Johnson 1963:553-554).

Graffam concludes his article by saying, "It may be the first time in educational history, in fact, that a college president or dean can provide scientifically valid answers to questions by anxious parents who inquire what the college will do to their sons and daughters besides increasing their store of knowledge" (Graffam 1970:5).

The 16PF is probably the best personality questionnaire available for research purposes in education. It surpasses other tests in the number of traits measured and in the longitudinal developmental research investigating personality dimensions.

"One substantial advantage of the factored ability of personality tests . . . is that the criterion relationships, the occupational profiles, the clinical diagnostic weights, . . . which are built up at such a cost of human effort, do not have to be thrown away when the test is scrapped and revised" (Cattell 1968:324). The utility of the 16PF is an important factor to consider. Even if the music teacher profiles are established on a test which is later revised, the fact that the new test better measures the personality dimensions does not render the previous profile data invalid. In fact, it strengthens it because of better measurement. Cross validation procedures can be used to assure the continued utility of the 16PF data.

THE 16 PRIMARY SOURCE TRAIT FACTORS

The 16 primary factors are shown in Table 1. "Some twenty-three primary personality factors have now been recognized (Cattell and Ford in preparation) in Q-data and also, more approximately, in L-data . . ." (Cattell and others 1970:7). The 16PF includes the 16 factors which account for the greatest variance in personality information.

Factor A - Sizothymia versus Affectothymia

In questionnaire responses, A+ individuals express marked preference for occupations dealing with people. They enjoy social recognition and are generally willing to go along with expediency.

The A- person likes things or words, works alone, takes hardheaded intellectual approaches, and tends to reject compromise. A+ persons are natural joiners more readily forming active groups and are more generous in personal relationships, less afraid of criticism, better able to remember names of people, but more less dependable in precision work and more casual in meeting obligations. "The highest ranking A+ occupations in scores are social workers and business executives and the lowest in A- scores are artists, electricians, and research scientists" (Cattell and others 1970:81).

Musicians tend towards a moderate A- score. Teachers also appear A- while this factor was not significant for university professors (Cattell and others 1970:209).

Factor B - Low IQ versus High IQ

While this factor will not correlate as highly with the usual speeded intelligence test, "There is probably little justification in consuming more time in assessing intelligence, relative to other factors, than is given by this brief B scale" (Cattell and others 1970:83). It is apparent that general ability is important in individual differences. ". . . the main dimension of general mental capacity is always important and accordingly is here given its proper role alongside its poor dimensions" (Cattell and others 1970:82). A general IQ factor can be estimated from the sten score for an intelligence quotient figure.

Factor B is not significant in studies with musicians. Throughout the educator samples, a B+ score is significant. Both teachers and musicians tend to be more intelligent than samples from other occupations and the general public.

Factor C - Emotional Instability or Ego Weakness versus Higher Ego Strength

In the questionnaire data, the C- person is easily annoyed by things and people, is dissatisfied with the world situation, his family, the restrictions of life, and his own health; and he feels unable to cope with life. He tends to show general neurotic responses.

C+ individuals are more frequently leaders. Clinically, most disorders show low ego strength scores. "It is the most general pathological contributor being found in neuroticism, psychotics, alcoholics, drug addicts, and so forth" (Cattell and others 1970:84). Individuals such as airline pilots tend to be high in Factor C while postmen, clerks, janitors, and writers are lower than average.

Musicians tend toward C- scores. This factor is C+ for university professors but has mixed results for other educators. Studies in Britain and the

United States with student teachers show a C- tendency (Cattell and others 1970:168-169).

Factor E - Submissiveness versus Dominance

This factor has different patterns for males and females. In females, the dominance traits such as socially poised and prominent are more highly loaded in the E factor than they are for males.

Dominance tends to be correlated positively to social status and not substantially with leadership. It is most associated with occupations requiring boldness and courage. "Dominance is negatively related to school achievement at all ages up to graduate university work, since docility seems to enhance examination performance" (Cattell and others 1970:86).

Factor E is a broad, temperamental, dispositional personality trait which, though showing a low but significant correlation with the two traits of need achievement measured by the MAT test (self-assertion erg and sentiment to the self) is structurally quite distinct from them. Occupationally, higher dominance scores go to farmers, cooks, and janitors. "It is appreciably influenced by heredity and is one of the personality factors distinguishing the sexes" (Cattell and others 1970:86-87). Musicians tend to be high in Factor E as do teachers at all levels.

Factor F - Desurgency versus Surgency

This is one of the most important components in extraversion. Surgent persons have generally had an easier, less punishing, more optimism-creating environment and are more happy-go-lucky. "Significant associations have been found sociologically between surgency and bachelorhood . . ." (Cattell and others 1970.87).

Desurgency is not to be confused with depression. It is soberness and caution. Elected leaders are far higher than followers in surgency but effective leaders show scarce significance. Airline hostesses and sales managers are high while artists, university professors and administrators and physicians are low.

It appears that musicians are high in surgency while educators are low. Student teachers are significantly low.

Factor H - Threctia versus Parmia

"This is a strongly defined factor which appears with high persistence even in rough factorings of ratings and in questionnaires" (Cattell and others 1970:91). The H- individual tends to be shy, tormented by an unreasonable sense of inferiority, slow and impeded in expressing himself, disliking occupations with personal contacts, preferring one or two close friends to large groups, and not able to keep in contact with all that is going on around him. "Present evidence indicates it to be one of the two or three most highly inherited of personality factors" (Cattell and others 1970:92).

The H+ person shows little inhibition by environmental threat and is rated lazy in childhood and thickskinned in social interaction. This constitutional insusceptibility to inhibition in turn generates the boldness in social, sexual, emotional, and physical danger situations which comes out in H+ individuals.

Factor H is high in airmen and administrators and low in farmers. This factor seems to be an important one in distinguishing suitability for those occupations demanding ability to face wear and tear in dealing with people and grueling emotional situations (Cattell and others 1970:92).

H+ individuals have a history of being more involved in organizing clubs or teams. Musicians are H+ as are student teachers and university professors.

Factor I - Harria versus Premsia

The I+ person has a fastidious dislike of crude people and rough occupations, a romantic liking for travel and new experience, a love of dramatics, a labile, somewhat unrealistic, imaginative, aesthetic mind, and a certain impracticality in general affairs (Cattell and others 1970:93). There are indications that artists score high on this factor and that females are decidedly higher than males, neurotics higher than normals.

This may be the reason that musicians score high on the Feminine scales of other personality tests such as the MMPI. The I+ person receives more descriptions as fussing, slowing up group performance in arriving at decisions, and making negative socioemotional remarks. The I- person represents some sort of tough, masculine, practical, mature group solidarity-generating and realistic (no nonsense) temperamental dimension (Cattell and others 1970:93).

Premsia (I+) is associated primarily with an overprotected or sheltering from urgent demands of life upbringing. "Increasing evidence points to its being clinically the matrix of attitudes out of which neurotic maladjustment can arise, especially the syndrones of conversion hysteria and hypochondria" (Cattell and others 1970:93). It is associated with sociopathic and drug addiction behavior. I+ is associated with student smokers while I- with nonsmokers.

Employment counselors, musicians, psychiatric technicians, artists, writers, and painters are high in Premsia. Factor I does not appear to be a factor for educators which should be of interest in this proposed investigation. Premsia tends to be significantly associated with mental breakdown both psychotic and neurotic (Cattell and others 1970:95).

Factor L - Alaxia versus Protension

Protension refers to projection and inner tension. High protension is noticeable in creative fields and has an important role in paranoia. Alaxia refers to persons who are easy going, friendly, and may lack ambition and striving. The protensive person comes from a parental home which he admired, and had lively intellectual interests. He is contemptuous and scrupulously correct in behavior while being annoyed at other persons putting on airs. He is skeptical of alleged idealistic motives in others. He lists a relatively high number of annoyances, is uninfluenced by the views of prominent people, and shows a high inner tension which takes the form of a feeling of social insecurity. There is some suggestion that protension is a preferred method of handling anxiety among intelligent persons.

The protensive person may be rated unpopular. School counselors and social workers run low in protension while time study engineers and social workers are L+. Musicians are L+ while educators including public school teachers, student teachers, and university professors are L-.

Factor M - Praxernia versus Autia

The M+ person has an intense subjectivity and inner mental life. "Although carried forward on inner tides of confidence, and definitely inclined to be disregardful of practical manners, he actually has higher internal, spasmodic anxiety and conflict tensions than the praxernic person, walks and talks in his sleep, and alternates periods of placid disregard of practical considerations with hysteroid episodes of relatively immature demanding and overwrought behavior" (Cattell and others 1970:98). Praxernia conveys the serious, practical concern with awkward details while Autia is related to internally autonomous thinking.

High M individuals tend to feel unaccepted in groups. Artists, researchers, some planning executives, and editors are high in M. ". . . high M has been found most significantly to distinguish the more creative researchers and artists from administrators and teachers of the same eminence" (Cattell and others 1970:99).

Musicians are definitely M+ as are teachers.

Factor N - Naivete versus Shrewdness

"This factor is not so well established in behavior ratings as in the questionnaire . . . it has still to be pursued to higher validity in the questionnaire realm" (Cattell and others 1970:99). High N indicates a person who is ingenious, sharp at clinical diagnosis, flexible in viewpoint, inclined to study the angles, alert to manners, to social obligations, and to social reactions of others. "The pattern represents some form of intellectual-educational development in terms of shrewd tactical skills, but there is little danger of its being confused with intelligence, though it does correlate positively both with intelligence and dominance" (Cattell and others 1970:100). Low N refers to almost naive emotional genuineness, complete directness, and spontaneous outspokenness.

High N persons are found in the skilled professions and precision occupations such as accountants, engineers, and electricians while Low N groups are missionary priests, nurses, psychiatric technicians, cooks, and convicts. Hadley found high N correlated with teaching success and there are other indications that low N individuals are more trusted and liked -- perhaps by children (Cattell and others 1970:100).

N does not appear significant for musicians. Only one study disagreed with the previously mentioned results and found industrial arts teachers to be N+ (Cattell and others 1970:169).

Factor O - Untroubled Adequacy versus Guilt Proneness

The O+ person feels he is unstable, reports overfatigue from exciting situation, is unable to sleep through worrying, feels inadequate to meet the rough daily demands of life, is easily downhearted and remorseful, feels that people are not as moral as they should be, is inclinded to piety, prefers books and quiet interests to people and noise. He shows a mixture of hypochondriacal and neurasthenic symptoms but with phobias and anxieties most prominent (Cattell and others 1970:102).

High scores occur in certain religious groups, artists, farmers, and editorial workers. Low scores occur with professional athletes, electricians, mechanics, nurses, and sales managers. Guilt proneness is one of the largest factors in anxiety and is high in neurotics, alcoholics, and psychotics. High O is strongly weighted against leadership in face to face situations and against success in tasks with sudden emergencies. Both musicians and teacher groups appear low in O.

Factor Q1 - Conservatism of Temperament versus Radicalism

This is the first of the factors which appears in questionnaire data but has not been isolated in behavior ratings. Q1 is more than a mere set of acquired radical, intellectual, political, and religious attitudes.

Q1+ persons are more well informed, more inclined to experiment with problem solutions, less inclined to moralize, and less unquestionning about views generally. While adolescents are high in need to revolt, Q1 is actually higher in people in middle life. Q1+ persons express more interest in science than religion, more interest in analytical thought, modern essays, and reading as opposed to class instruction, in breaking the crust of custom and tradition, and in leading and persuading people.

Q1+ is high in executives, university professors, and especially scientific researchers. It is low in policemen, nurses, semiskilled and unskilled groups. Musicians are definitely high in Q1 as are some student teachers and university professors. There are mixed results with other teacher populations.

Q2 - Group Dependency versus Self Sufficiency

The Q2+ person is resolute and accustomed to making his own decisions while the Q2- person goes along with the group, definitely depends on social approval, and is conventional and fashionable. Q2 is high for farmers, writers, scientists, and ciminals. The high Q2 person is dissatisfied with group integration, makes remarks which are more frequently independent solutions than questions and tends to be rejected.

Q2+ persons are associated with self sufficiency leading to significant association with scholastic success and coronary attacks. Q2+ is a significant factor with musicians but does not appear for educators. Music therapists tend to be low in this factor.

Q3 - Low Self Sentiment Integration versus High Strength of Self Sentiment

Cattell's major hypothesis with factor Q3 is that it represents the strength of the individual's concern about his self concept and social image. The Q3+ person shows socially approved character responses, self control, persistence, foresight, considerateness of others, conscientiousness, and regard for etiquette and social reputation. High Q3 is associated with leaders and success in mathematical, mechanical, and productive organizational activities. It is high in airline pilots, university administrators, electricians, and psychiatric technicians and is related to success in school (Cattell and others 1970:106).

The Q3+ person has crystallized for himself a clear, consistent, admired pattern of socially approved behavior, to which he makes definite efforts to conform (Cattell and others 1970:107). What is being measured here is the amount of concern about and regard for these standards. Both teachers and musicians are high in Q3.

Factor Q4 - Low Ergic Tension versus High Ergic Tension

Ergic tension shows itself by the individual's being irrationally worried, tense, irritable, anxious, and in turmoil. Persons high in Q4 seldom achieve leadership. Editors and persons in jobs which give little self expression but release an onslaught of environmental demands run high while airline pilots and hostesses are low.

Q4- distinguishes the overachievers from the underachievers of the same intelligence level. In groups where high activity levels are required, performance has been found positively related to Q4. "For this reason, such terms as 'id pressure' or 'frustrated drive' -- or, more exactly, the present term ergic tension (in quantitative dynamic concepts) -- express it better than such former terms and metaphors as 'tensions'" (Cattell and others 1970:109).

Motivation Factors

"It is commonplace among experienced observers in almost any field of endeavor that personality and ability together decide the outcome, their relative importance varying according to the form of achievement involved" (Cattell 1965:309). Personality includes motivation and value variables.

The educational psychologist has primarily attempted to predict achievement from abilities alone. Until a few years ago, this might have been a proper approach considering the lack of valid personality and motivation tests. This is no longer true.

From research conducted by Cattell and others, personality factors apparently contribute to school achievement about as much as abilities (Cattell 1965:310). "Our evidence here is largely on various ergic tension levels and on sentiment structure development in relation to school achievement" (Cattell 1965:313). When persons are at a high anxiety level, greater effort may actually lead to reduced achievement.

The importance of discussing achievement is crucial. It matters little whether achievement refers to pupil learning in school or to teachers achieving in the classroom.

Cattell states that the percentage of the variance in a criterion accounted for by some personality measure which predicts it (i.e. achievement), is the square of their correlation, not just the coefficient itself. The correlation typically found between an intelligence test and school achievement is 0.5.

"This means that $(0.5)^2$ x 100, i.e. 25 percent of the range observed in achievement is associated with the intelligence variability in the particular population" (Cattell 1965:313). If the intelligence differences were eliminated, the variance in achievement would be 75 percent.

Thurstone, by including several primary abilities with intelligence, was able to provide for 25 to 30 percent of the variance. "The experimental evidence on motivation measures, with MAT is too recent for any precise conclusion to be drawn until checked, but it looks as if about 20 percent will be found associated with motivation" (Cattell 1965:314). Cattell claims that with three batteries of tests (IPAT, Culture-Fair IQ Test, 16PF, and MAT), 60 to 80 percent of the variance in achievement can be estimated.

Motivation Analysis Test (MAT)

For over 20 years, Cattell and associates have been investigating motivation and isolated the separate factors and components. "Our interests, attitudes, and motivational habits are not a formless, chaotic aggregate, but show a characteristic, discoverable 'simple structure,' in which certain major common traits repeatedly stand out, even among people of different ages and different cultural groups" (Cattell and Butcher 1968:73). It is as meaningful to measure these individual strengths of major motivation sources in the dynamic realm as it is to measure abilities and temperamental dimensions.

When the nature of motivational factors is examined, some appear to be drives. These factors are called ergs.

Other factors differ from ergs (drives) in that their contributory attitudes have some cultural object in common that satisfy quite a variety of emotions. "It is hypothesized that ergs are innate and common to all mankind, whereas the sentiment factors represent learned groupings of attitudes and would be expected to differ in number and nature from culture to culture, although this has not yet been proved" (Cattell and Butcher 1968:74). The ergs are to be considered sources of energy or reactivity, and the current strengths of energy in an individual at a given time will be represented by his various ergic tension levels. The sentiments reflect the social institutions of the culture by which the individual is influenced in his years of learning. Sentiments include one's career, religion, home, hobbies, recreations, and so forth. The measurement of the strength of a sentiment provides an indication of the interest the individual has invested in the given concrete or abstract object. Common dynamic traits can be measured.

The ergs measured on the MAT test are: (1) mating, (2) assertiveness, (3) fear, (4) narcism-comfort, and (5) pugnacity-sadism. The sentiments are: (6) sentiment to self concept including the social reputation component of the self and control and understanding of the self, per se, (7) the superego sentiment to socio parental sanctions, the moral-ethical ideal self, (8) career - profession, (9) sweetheart - spouse, and (10) home - parental.

> The ten dynamic structures in MAT were chosen carefully to give the most dynamically, clinically-useful measures among the roughly twenty dynamic factors which research to date has established to be representative, and comprehensive in coverage, of adult motivations . . . The choice is not arbitrary, but comes from an overview of many years of correlational and factor analytic research . . . From among the known dozen or so sentiment factors, the five in MAT, . . . are chosen as the major interest attachments common to most people and most relevant to clinical, educational, and occupational understanding of their lives (Cattell and Horn 1964:2).

Reliability and Validity

The reliability or consistency has been measured by short term test - retest and homogeneity coefficients. Validity was established by the correlation of the MAT scales to items in the development study. The concept validity is estimated by correlating each of the ten scale scores with the corresponding pure factor. This is achieved from the matrices giving the correlations among the subtests, and by using multiple correlation and least squares procedures to estimate the correlation of battery and factor. "The multiple R's so obtained show validities for all ten factors which lie in the .90's and so need no individual discussion" (Cattell and Horn 1964:33). Taking into consideration the short time and item length of any one of the ten scales, these concept validities are fairly high.

Because dynamic traits will change in time, a truly valid dynamic measurement instrument implies that a factor is expected to change due to circumstantial stimuli and internal conditions. A high reliability or stability coefficient would reveal a poor and insensitive motivation test according to Cattell's argument. The stability for the sentiments is higher than for ergic tensions. Cattell says:

> In evaluating and using the reliabilities and validities for the single factors here, the psychologist will recognize that comparisons with the face values of simple intelligence and achievement tests are either pointless or misleading. By excluding ability variance from motivation measures, one necessarily reduces the apparent reliability, just as the reliability of ability measures, i.e. of the true factor measures, would appear much lower if motivation were partialled out. The values to which readers are accustomed in traditional tests are, indeed, incorrect ones, inflated by vehicle variance, response sets, and instrument factors. Considering these more sophisticated issues, and the brevity of a measure which is less than a tenth of a test with less than an hour of actual testing, the validities and reliabilities here are perhaps as high as they will go for measures of this kind (Cattell and Horn 1964:35).

Ergs

1) Mating is the sex drive and determines interest in the opposite sex, romantic literature, courting behavior, beauty, and sensuality.

2) Assertiveness is one of the roots of ambition and status seeking and is similar to pride but best called vanity or assertiveness. It shows itself in striving for admiration, wishing to excel in competition, aiming to be handsomely dressed, and moving in high circles.

3) Fear or escape refer to security. There is interest in avoiding illness, accident, loss of financial security, military threats, and death.

TABLE 2.--MAT FACTORS

Dynamic Factor	1. Dependability Coefficients	2. Alpha	3. Arbitrary Split-Half	4. Stability Coefficients	5. Validity
Fear	.70	.58	.54	.48	.60
Mating	.66	.45	.49	.51	.69
Assertiveness	.64	.33	.47	.53	.53
Narcism	.60	.43	.43	.53	.52
Pugnacity	.51	.39	.33	.41	.72
Career	.53	.42	.37	.39	.66
Sweetheart	.74	.63	.58	.47	.76
Home-Parental	.81	.65	.70	.65	.66
Superego	.67	.44	.50	.46	.61
Self Sentiment	.78	.71	.64	.69	.76

1 Based on N=155 Air Force enlisted men.

2 Based on N=227 adults in various walks of life.

3 Based on N=151 college students

4 Based on over a five week period, N=101 college students

5 Based on the correlation (corrected for overlap) between primary factor scores overall items in the development study and the factor measured by the items actually selected for the test. (Cattell and Horn 1964:5).

Circumstantial validity (the extent to which the test for X and the concept for X has correlations with other factors, A, B, C, D, etc.) has not been calculated due to the need for new data involving a wide spectrum of general personality, ability, and motivational variables.

"Strictly, it should be called the escape erg, with fear as the accompanying emotion" (Cattell and Horn 1964:22).

4) Narcism-comfort refers to Freud's narcistic, sex component, It is directed to sensual indulgences of all kinds such as food and smoking, to ease, self love, and avoidance of onerous duties.

5) Pugnacity-sadism erg is the drive to attack, damage, inflict pain, and destroy. Its accompanying emotion is rage in the extreme.

Sentiments

Because the sentiments are widely known, Cattell does not describe them in detail in the Handbook. "In predicting what an adult will do, however, it is evidently quite as important to know the motivation available from these acquired sentiments as from his ergs" (Cattell and Horn 1964:22). The evaluation of a man's potential in a career should include measures of the strength of his interest in the career as well as his degree of concern about his general reputation and self concept, especially of the level of dependability implied by his level of superego development.

The sentiments show how far a person's interest in a particular course of action is likely to be affected by his existing stable emotional investments in particular objects. The sentiment measures represent degrees of development of acquired dynamic structures through personality learning.

Predictive Relationships and Interrelationships (Subproblem 2)

This subproblem referred to the relationships and interrelationships of the psychological factors to the music teaching success criteria. Male and female music educator samples were divided into three groups for some analyses after being rank ordered for the gain and rating criteria: successful, average, and unsuccessful. Because of possible sex differences on certain variables, the relationships and interrelationships between the 9 success groups, sex, and the 36 personality variables were first investigated. A 9 by 2 by 36 factorial design using the multivariate analysis of variance technique (MANOVA) was conducted.

Multiple regression equations were formulated for predicting music teaching success in terms of G (Gain), R (Rating), G-R (Gain minus Rating), and G+R (Gain plus Rating). The equations derived for the derivation sample were cross validated.

To further investigate predictive relationships and interrelationships of the personality and motivational variables, multiple regression equations were also formulated for subgroups of music educators (female choral, male choral, female instrumental and male instrumental) and cross validated to the music student teacher sample. Because of the small number of subjects, the total music educator sample was used.

The participants were tested individually and in groups using both forms of the 16PF and Form A of the Motivation Analysis Test. All personality testing was completed in April, 1972.

Profiles (Subproblem 3)

The third subproblem referred to the differences in personality profiles for successful and unsuccessful male and female music educators. The male and female samples were stratified into 3 success groups for both the G and R criteria. The upper and lower 25 percent were classified successful and unsuccessful respectively. Two group discriminant function analyses investigated profile differences for the G and R criteria for successful and unsuccessful male and female music educators. The following comparisons were investigated for each criterion: (1) successful males and successful females, (2) unsuccessful males and unsuccessful females, (3) successful males and unsuccessful males, and (4) successful females and unsuccessful females. These analyses are similar to multiple regression analyses with dummy variables. These profiles helped to describe the difference in personality and motivation.

Implications for Counseling and Advising (Subproblem 4)

Subproblem four refers to the practical applications to which the findings may be used. While it was not a researchable problem per se, it did provide an opportunity to compare the findings of this study with other investigations and to offer recommendations to teacher training institutions for incorporating the use of personality and motivational information in advising and counseling procedures. "To the field of music education, the ultimate value in research is realized when it is successfully applied to some element of the teaching-learning process" (Schneider and Cady 1965:20) Once the psychometric profiles were obtained, inferences from these may aid in counseling and advising undergraduate music education students. Discriminant and regression analyses provide individual variable factor weights which can be incorporated in statistical procedures for estimating potential success of persons not presently in the music education profession, e.g. undergraduate music education student teachers. These techniques provide the way in which college advisers can obtain data from their students and estimate potential music teaching success from personality variables alone. Two cross validation groups were offered in this investigation: the music student teacher sample and 20 percent of the music educator sample. If significant equations for predicting success relationships exist between personality and motivational variables and criteria and are significantly cross validated, it can be shown that inferences can be made for using this information in the undergraduate music teacher training programs.

The intent of this research project was not to propose the thesis that personality information should be a primary concern in the selection and retention of potential music educators. Rather, personality and motivational information can be incorporated in programs of behavior modification to assist college students to overcome difficulties before these are faced during student teaching or on the job. For example, many persons have a fear of actually teaching children; yet, they have a strong desire to teach music. Obvious implications would refer to implementing a program of gradual experiences with children during the college years which will modify that fear erg.

> There has been much interest in recent years in behavior modification in education settings. Although the work is still relatively limited, we can look forward in future years to a sizable increase in activity in this area. Moreover, we will see an increase in precision and an increase in the use of instrumentation (Holland and Doran 1973:307).

While in music education behavior modification per se may not be the answer, structuring the teacher training experiences certainly will help persons with potential personality and motivational problems in certain areas.

> . . . when others read the results of a study of teacher effectiveness or school effectiveness, their confidence in the results may depend largely upon their confidence in the judgment of the particular researcher for the study, and not just the reported results themselves. More specifically, their confidence may be based upon how well, or how meaningfully, some of the variables in the study were scaled (Light 1973:319).

Morgan and Morgan state that a poor personal approach to music teaching always proves to be a handicap and that these items should not be neglected in training.

> As intangible as desirable personality may be, many of the component parts are very recognizable and it has been proved beyond a doubt that many phases of a desirable personality can be developed (Morgan and Morgan 1960:99).

While it has not been proved beyond a doubt, certainly some behaviors, through specific clinical and psychometric techniques, can be modifiable by training.

Moses emphasizes the importance of personality:

> A pleasing personality is one of the greatest assets that a [music] teacher can develop. It is to be hoped that teacher-training institutions and inservice training courses will spend more time in the development of teacher personality (Moses 1970:185).

REFERENCES

Albert, H. R., "An Analysis of Teacher Rating by Pupils in San Antonio, Texas," Educational Administration and Supervision, XVI, January, 1941, 267-274.

Aleamoni, Lawrence M., "Illinois Course Evaluation Questionnaire (CEQ) Results Interpretation Manual Form 66 and Form 32," Research Report #331, 1972.

Aleamoni, Lawrence M., "A Review of Recent Reliability and Validity Studies on the Illinois Course Evaluation Questionnaire (CEQ)," Research Memorandum #127, May 1972.

Anderson, James Harold, "The Predictive Value of Selected Factors for Student Teacher Success at Western Carolina University," unpublished doctoral dissertation, University of Tennessee, 1968.

Anderson, John Martin, "The Use of Musical Talent, Personality, and Vocational Interest Factors in Predicting Success for Student Music Teachers," unpublished doctoral dissertation, University of Southern California, 1965.

Ausubel, David P., Educational Psychology, New York: Holt, Rinehart and Winston, 1968.

Barr, A. S., William H. Burton, and Lee J. Brueckner, Supervision: Prinicples and Practices in the Improvement of Instruction, New York: D. Appleton Century Company, 1938.

Barth, George William, "Some Personality and Temperament Characteristics of Selected Music Teachers," unpublished doctoral dissertation, University of Southern California, 1961.

Bereiter, Carl E., "Some Persisting Dilemmas in the Measurement of Change," Problems in Measuring Change, Chester W. Harris, ed., Madison: University of Wisconsin Press, 1963.

Betts, Gilbert L., "Pupil Achievement and the NS Trait in Teachers," in The Measurement of Teaching Efficiency, Helen M. Walker, ed., New York: The Macmillan Company, 1935, 145-203.

Bodenstab, Ernest W., "Personality Patterns and Heterogeneity Among Teaching Interest Areas: A Multivariate Approach," AERA Paper Abstracts, 1968 Annual Meeting, 319.

Borkowski, Francis Thomas, "The Relationship of Quality of Work in Undergraduate Music Curricula to Effectiveness of Instrumental Music Teachers in the Public Schools," unpublished doctoral dissertation, West Virginia University, 1967.

Brophy, Jere E., "Stability of Teacher Effectiveness," American Educational Research Journal, Summer 1973.

Brown, Robert E., "A Study Concerning the Major Weaknesses of Music Teachers in the Secondary Schools of Kansas," unpublished master's thesis, University of Kansas, 1955.

Cady, Henry L., "A Conference on Research in Music Education," Council for Research in Music Education Bulletin, 18, Fall 1969, 10-21.

Cass, James, "Accountable To Whom? For What?" Saturday Review, March 20, 1971, 41.

Cattell, Raymond B., "Personality Measurement Functionally Related To Source Trait Structures," Measurement in Personality and Cognition, Messick & Ross, eds., New York: Wiley, 1962.

Cattell, Raymond B., The Scientific Analysis of Personality, Baltimore; Penquin, 1965.

Cattell, Raymond B., and J. J. Butcher, The Prediction of Achievement and Creativity, Indianapolis: Bobbs-Merrill, 1968.

Cattell, Raymond B., Herbert W. Eber, and Maurice M. Tatsuoka, Handbook for the Sixteen Personality Factor Questionnaire (16PF), Champaign: Institute for Personality and Ability Testing, 1970.

Cattell, Raymond B., and John L. Horn, Handbook for the Motivation Analysis Test (MAT), Champaign: Institute for Personality and Ability Testing, 1964.

Centra, John A., "Do Student Ratings of Teachers Improve Instruction?" Change Magazine, April 1973, 12-13.

Charters, W. W. Jr., "Survival in the Profession: A Criterion For Selecting Teacher Trainees," Journal of Teacher Education, 7:253, September 1956.

Colwell, Richard J., Interpretive Manual - Music Achievement Tests 1 and 2, Chicago: Follett, 1969.

Cronbach, Lee J. and Lita Furby, "How We Should Measure 'Change'--Or Should We?" Psychological Bulletin, 1970, 74:68-80.

Dayton, C. Mitchell, "The t-Test: Parametric or Non-Parametric," 1967 AERA Annual Report.

Dewey, John, Democracy and Education, New York: Free Press, 1916.

Dewitt, Charles Jay, "A Study of Selected Variables in Discriminating Between Contrasting Levels of Student Teaching Performance," unpublished doctoral dissertation, University of Virginia, 1970.

Duda, Walter Boleslav, "The Prediction of Three Major Dimensions of Teacher Behavior for Student Teachers in Music Education," unpublished doctoral dissertation, University of Illinois, 1961.

Dyer, H. S., R. L. Linn, and M. J. Patton, "A Comparison of Four Methods of Obtaining Discrepancy Measures Based on Observed and Predicted School System Means on Achievement Tests," American Educational Research Journal, 1969, 6.

Ehlert, Jackson K., "Causes for Failure Among Music Teachers," Music Educators Journal, 37, 3, January 1951, 36-38.

Ernest, David J., "The Prediction of Academic Success of College Music Majors," Journal of Research in Music Education, XVIII, 3, Fall 1970, 273-276.

Fosse, John B., Critique - "The Relationship of Pre-Service Measures to Ratings of Music Teaching Success," Charles H. Benner. Council for Research in Music Education Bulletin, 7, Spring 1966.

Fosse, John Benjamin, "The Prediction of Teaching Effectiveness: An Investigation of the Relationships Among High School Band Contest Ratings, Teacher Characteristics, and School Environment Factors," unpublished doctoral dissertation, Northwestern University, 1965.

George, Warren Edwin, "Significant Predictors for College Achievement in Specified Areas of Music Education and Identification of Potential Graduates (Volumes I and II)," unpublished doctoral dissertation, University of Kansas, 1969.

Getzels, J. W. and P. W. Jackson, "The Teacher's Personality and Characteristics," Handbook of Research on Teaching, N. L. Gage, ed., Chicago: Rand McNally, 1963, 506-582.

Glass, Gene V. and Julian C. Stanley, Statistical Methods in Education and Psychology, Englewood Cliffs: Prentice-Hall, Inc., 1970.

Gordon, Edwin, "Third Year Results of a Five Year Longitudinal Study of the Musical Achievement of Culturally Disadvantaged Students," in Experimental Research in the Psychology of Music: 8, Edwin Gordon, ed., Iowa City: University of Iowa Press, 1972, 45-57.

Graffam, Donald T., Dickinson College Changes Personality, Dickinson College, 1970.

Guilford, J. P., Psychometric Methods, New York: McGraw Hill, 1954.

Heil, Louis M., and Marion Powell and Irwin Feifer, Characteristics of Teacher Behavior and Competency Related to the Achievement of Different Kinds of Children in Several Elementary Grades, Washington: U. S. Office of Education, Cooperative Research, SAE 7285, May 1960.

Hellfritzsch, A. G., "A Factor Analysis of Teacher Abilities," Journal of Experimental Education, XIV, December, 1945.

Hill, John D., "A Study of the Musical Achievement of Culturally Deprived Children and Culturally Advantaged Children at the Elementary School Level," in Experimental Research in the Psychology of Music VI, Edwin Gordon, ed., Iowa City: University of Iowa Press, 1970, 95-123.

Hoffer, Charles R., Teaching Music in the Secondary Schools - Second Edition, Belmont: Wadsworth Publishing Company Inc., 1973.

Holland, James G. and Judith Doran, "Instrumentation of Research in Teaching," Second Handbook of Research on Teaching, ed. Robert M. W. Travers, Chicago: Rand McNally, Inc., 1973, 286-317.

Klotman, Robert H., The School Music Administrator and Supervisor: Catalysts for Change in Music Education, Englewood Cliffs: Prentice-Hall, Inc., 1973.

Lehman, Paul R., The Seventh Mental Measurements Yearbook, Volume I, ed. Oscar K. Buros, Highland Park: The Gryphon Press, 1972, 247-249.

Lehman, Paul R., Tests and Measurements in Music, Prentice-Hall, Inc., 1968.

Light, Richard J., "Issues In The Analysis of Qualitative Data," Second Handbook of Research on Teaching, ed. Robert M. W. Travers, Chicago: Rand McNaily and Co.,318-381.

Lord, Frederic N., "Elementary Models for Measuring Change," Problems in Measuring Change, Chester W. Harris, ed., Madison: University of Wisconsin Press, 1963.

Lord, Frederic N., "Statistical Adjustments When Comparing Pre-existing Groups," Psychological Bulletin, 1972.

Lucio, William H. and John D. McNeil, Supervision: A Synthesis of Thought and Action - Second Edition, New York: McGraw Hill, 1969.

Lutz, Warren William, "The Personality Characteristics and Experiential Backgrounds of Successful High School Instrumental Music Teachers," unpublished doctoral dissertation, University of Illinois, 1963.

Mahan, J. Maurice, Manual of Interpretation for the Illinois Teacher Evaluation Questionnaire, Champaign: Measurement and Research Division, Office of Instructional Resources, June 1970.

Maier, John Arthur, "The Relationship of Selected Factors to Academic and Musical Success in the Two Year, Basic Music Theory Curriculum at the State University of New York - Fredonia, New York," unpublished doctoral dissertation, The Ohio State University, 1970.

McCall, William A., How To Measure In Education, New York: The Macmillan Company, 1972.

Mitzel, Harold E., "Teacher Effectiveness," Encyclopedia of Educational Research, Third Edition, ed. Chester W. Harris, New York: The Macmillan Company, 1960, 1481.

Music Educators National Conference, Teacher Education in Music: Final Report, Washington: Music Educators National Conference, 1972, 9.

Music Educators Journal, May 1973, "NEA Reports."

Music Educators National Conference, 48.

Morgan, Hazel Nohavec, "The Ability Pattern of Senior College Students Majoring in Music Education at the University of Minnesota," unpublished doctoral dissertation, University of Minnesota, 1946.

Morgan, Russell Van Dyke and Hazel Nohavec Morgan, Music Education in Action, Park Ridge: Neil A. Kjos Music Company Publishers, 1960.

Morsh, Joseph E., et al, "Student Achievement as a Measure of Instructor Effectiveness," Journal of Educational Psychology, XLVII, February, 1956.

Moses, Harry E., Developing and Administering a Comprehensive High School Music Program, West Nyack: Parker Publishing Company, Inc., 1970.

National Council of State Supervisors of Music, Guidelines in Music Education: Supportive Requirements, Washington: Music Educators National Conference, 1972, 9.

Nunnally, Jim, "The Analysis of Profile Data," Psychological Bulletin, 1962, 59, 4, 311-319.

O'Connor, Edward F., Jr., "Extending Classical Test Theory to the Measurement of Change," Review of Educational Research, 42, 1, 1972.

Picerno, Vincent Joseph, "Personal Characteristics of Some Successful Music Teachers in Erie County, New York," unpublished doctoral dissertation, Northwestern University, 1956.

Poland, William, "Measuring Musical Behavior and Predicting Success in the University School of Music," AERA Annual Meeting Presentation, February 14, 1963, Chicago, The Ohio State University.

Ryans, David G., Characteristics of Teachers, Washington: American Council on Education, 1960, 1.

Schneider, Erwin H. and Henry L. Cady, Evaluation and Synthesis of Research Studies Relating to Music Education, Washington: U. S. Office of Education, Cooperative Research, E-016, The Ohio State University, 1965.

Shanks, James Louis, "Concept Achievement in Science and Its Relationship to Some Non-Intellectual Characteristics of Prospective Elementary Teachers," unpublished doctoral dissertation, University of California at Berkeley, 1969.

Shatin, Leo, "Personality Traits of Music Therapists," Psychological Reports, 1968, 23.

Silberman, Charles E., Crisis in the Classroom: The Remaking of American Education, New York: Random House, 1970.

Simpson, Edwin L., Critique - "The Relationship of Quality of Work in Undergraduate Music Curricula to Effectiveness of Instrumental Music Teaching in the Public Schools," Francis Borkowski, Council for Research in Music Education Bulletin, 16, 30-33.

Smith, Martha Lucile, "A Study of Elementary Student Teacher Confidence in and Attitudes Toward Music and Changes that Occur in a Student Teaching Experience," unpublished doctoral dissertation, Michigan State University, 1969.

Tatsuoka, Maurice M., Discriminant Analysis: The Study of Group Differences, Champaign: Institute for Personality and Ability Testing, 1970.

Tatsuoka, Maurice M., Multivariate Analysis: Techniques for Educational and Psychological Research, New York: John Wiley and Sons, Inc., 1971.

Tatsuoka, Maurice M. and David V. Tiedeman, "Statistics as an Aspect of Scientific Method in Research on Teaching," Handbook of Research on Teaching, N. L. Gage, ed., Chicago: Rand McNally and Co., 1963, 142-170.

Whybrew, William, Measurement and Evaluation in Music - Second Edition, Dubuque: Wm. C. Brown Co., 1971.

Williams, Raymond Edmund, "The Measurement and Prediction of Cooperating Teacher Effectiveness in Music Education," unpublished doctoral dissertation, University of Illinois, 1958.

Wink, Richard L., "The Relationship of Self Concept and Selected Personality Variables to Achievement in Music Student Teaching," Journal of Research in Music Education, 3, Fall, 1970, 234-241.

CHAPTER II

A SURVEY OF RELATED LITERATURE

Introduction

There is a plethora of research studies in teacher education which relate to predicting success. Many have investigated achievement and other variables during college as preservice predictors of student teaching or music teaching success. These include: home, family, and demographic variables, experiential and knowledge variables, interest, motivation, and personality variables, and various teaching tasks or process variables. Often, relationships were sought which assisted not only in the prediction of potential teaching success but also aided in counseling and advising college majors. The most common approach has been to correlate variables to success criteria. Other studies have investigated groups, e.g. success groups, musical and nonmusical groups, and used statistical procedures to look for significant differences between the groups on the variables under investigation. In this manner, profiles were developed for predicting and counseling purposes.

There are various categories of research studies in the general area of predicting success. Music studies have investigated numerous variables related to musicianship at the elementary and secondary school levels as well as college. These studies are delimited from this survey of the literature except as they affect the criterion variables.

Investigations of the importance and effectiveness of various teaching methodologies and achievement variables relating to learning music and learning in general for potential success in college have been made. These have concentrated on college entrance variables and their relationship to successful achievement in college. The assumption is that success in college is related to teaching success. This appears to be valid because most states require a minimum of a bachelor's degree for certification in teaching. Before a teacher earns the opportunity to succeed or fail in teaching, he must first be certificated. Categories include: success in college, student teaching success, music teaching success, field surveys of music educators' opinions relating to the validity and usefulness of their college course experiences, and experts' opinions about the importance of selected preservice activities and experiences deemed crucial to music teaching success.

Outside of music, there are literally thousands of research investigations which attempted, in some manner, to predict teaching and student teaching success. Some of these are important and relate to music education and bear on the design of this study.

No one has developed an adequate model for predicting teaching success: this seems ironic considering the vast amount of research undertaken to investigate this area. Perhaps one reason for this fact is that no one has offered an acceptable theory of teaching success or what teaching is and ought to be. Instead, there are numerous theories with no one inclusive enough to cover all of the things that a teacher does in the course of the act or process called teaching. Research investigations have looked at specifics of the act, offered hypotheses, and rendered conclusions which may or may not generalize further than the immediate sample under investigation.

Not only is there a philosophic lag in teacher education and the teaching profession in general, but so too in the field of music education. Colwell emphasizes this philosophical problem:

> The genuine philosopher, trained to carefully scrutinize both the form and substance of any piece of thought, would find it difficult to attach much consistency or logic to the viewpoints variously followed by music educators. . . . The responsibility, therefore, finally devolves upon those who form the philosophy of the profession; they must state clearly the objectives they believe to be essential, delineating the areas of activity and the levels of achievement the child should experience in public school music. . . . Philosophy's responsibility must be first, to make statements of sufficient concreteness and specificity that they can be utilized for experimental research projects, to give direction to research; and second, to use the research now existing as a guide, if not a final answer, to problems of teaching and learning (Colwell 1967: 74-75).

Leonhard and House further emphasize the importance of philosophy to the music education profession:

> When we speak of a philosophy of music education, we refer to a system of basic beliefs which underlies and provides a basis for the operation of the musical enterprise in an educational setting. A philosophy should serve as a source of insight into the total music program and should assist music teachers in determining what the musical enterprise is all about, and how it should operate. A definitive philosophy is useful, even essential, for an operation as complex and as important as music education because concepts, theory, and practice rely on one another (Leonhard and House 1972:83-84).

Madsen and Madsen in discussing the importance of the relationship of philosophy to research states:

> Important relationships between "what is" and "what ought to be" merit extremely serious appraisal, as does the entire area of knowledge and how one knows anything. There is no substitute for coming directly or vicariously in contact with the most important ideas and positions available. To think without acting may result in loss, but to act without thinking may produce far worse consequences (Madsen and Madsen 1970:13).

Like Leonhard and House, Reimer believes that a philosophy for music education is found in aesthetic theory,-the nature and value of the arts:

> As this profession becomes more complex, it is reasonable to assume that differences of philosophy will also become more complex. The philosopher must accept and even approve of this condition. But, again, the fact that no "one true answer" is possible or desirable does not remove the need for reasoned, careful, systematic statements which help make the field more understandable to all who are involved with it. A philosophy, then, must be conceived as being "of a time," and must also give recognition to the fact that it can only provide a point of departure for practitioners of that time (Reimer 1970:2).

Yet, Reimer realizes that some philosophy, some underlying set of beliefs about the nature and value of music education, is absolutely necessary if one is to effectively and successfully practice his profession. "What the profession seems to need at the moment is not persuasion about any particular philosophy, but articulation, refinement, and careful application of the commonly held but largely unarticulated, unrefined, and imperfectly applied beliefs now current." (Reimer 1970:3)

Colwell, Madsen and Madsen, and Reimer highlight the dilemma of every researcher: conflicting ideas, philosophies, and beliefs must be resolved in the design of the research project. The philosophy accepted by the researcher constitutes his frame of reference and affect the kinds of questions he asks and the hypotheses he tests. Leonhard and House cogently urge the development and utilization of a philosophy by each and every music educator no matter at what level of professional practice he may be. This is essential because philosophy assists one to determine his "modus operandi." Yet, Colwell states that the music education profession as a whole does not ascribe to one philosophy or statement of beliefs, thus, creating a serious void for the music education researcher. Reimer emphasizes the need for philosophers to assist music educators to articulate clearly those objectives, experiences, and areas of activity that are important for children in public school music. The researcher, by choosing the design of his study, in essence, makes important philosophical decisions.

A researcher must thus attempt to: (1) accept one school of thought and devise a research design accordingly, or (2) allow for a pragmatic research design which includes as many important and crucial philosophical beliefs as possible. In either case, the philosophical base must be clear to the reader of the research report. This chapter attempts to provide the philosophical foundation for the design of this study.

Discussed first is the area of predicting success in college, student teaching, and teaching. Basic philosophical areas are related to the development of the subproblems of this investigation.

The second area, relationships among accountability criteria, highlights various evaluative schemes used in research. These generally fall into three categories of criterion or dependent variables: (1) product in terms of pupil growth, (2) product in terms of observed teacher behaviors, subjective ratings of the teacher by supervisors and others, and (3) process variables which are concerned with the relationships and interrelationships of the teaching act or setting, e.g. teacher-pupil, teacher-administrator, teacher-parent, environmental and other interactions of the classroom climate affecting the teaching process. Therefore, it was necessary to search the literature for assistance in the criterion domain for methodological design assistance.

The third area concentrates on studies which have used personality and motivation variables for classifying individuals into career categories and in predicting student teaching and teaching success in music. Highlighted are differences found in personality profiles for men and women. Special attention is paid to the specifics of the research designs and the methods used in determining significant findings.

SECTION ONE: PREDICTIVE MEASUREMENT OF SUCCESS

Since 1919 when a music aptitude test battery was developed by Carl Seashore, much effort has been expended in attempting to predict success with music for students at the elementary, secondary, and college levels. Tests by Seashore, Kwalwasser-Dykema, Drake, Wing, Bertley, and Gordon to mention a few have succeeded to some degree of usefulness in separating musical and nonmusical persons. Also, achievement tests by Knuth, Alice Synder Knuth, Aliferis,

Aliferis and Stecklin, Gordon, and Colwell have been used successfully within prediction programs for counseling and advising students.

Predicting Success in College

It is generally accepted that previous achievement is one of the best predictors of future achievement. Predictions for success in college can be made from a number of precollege variables.

Most colleges make some kinds of entrance decisions for freshman students based upon a combination of high school experience and college entrance test variables. The most common approach is to use the high school centile rank in combination with an American College Test score (ACT) or Scholastic Aptitude Test score (SAT). Those students with higher scores are predicted to have the best potential for success in college.

In music education, colleges may use some type of entrance test battery in music to predict various college course grades such as music theory, music history, applied music, and so forth. A number of studies have shown successfully that a combination of entering and testing variables are useful in predicting success for the college music major. A few of the best studies are highlighted below.

Ernest (1970) investigated the relationship between selected precollege scores and (a) college cumulative honor point and (b) cumulative honor point in music courses which were attained at the time of college graduation as a basis for predicting the degree of academic success of freshmen college students proposing a major in music. All available scores were obtained from the American College Tests (ACT), Minnesota Scholastic Aptitude Test (MSAT), Triggs Reading Survey, High School Rank (HSR), Cumulative College Honor Point Ratio (HPR), and Honor Point in Music Courses (MHP) for the 71 students who graduated from St. Cloud, Minnesota, State College, with a major in music between December, 1960, and August, 1967. No effort was made to establish data on those students entering college and not completing the bachelor's degree. He used the correlation between HPR and each precollege score and between MHP and each precollege score as well as formulating a multiple regression equation. The best single predictor for college cumulative honor point and honor point in music courses was high school (centile) rank (r=.43, 4=.44). A combination of HSR, Triggs, and MSAT produced the best multiple correlation coefficient (R=..574) as a predictor of college cumulative honor point or degree of academic success. Combining high school centile rank with Triggs as a predictor of the music honor point ratio had only a slightly better result (R=.449) than using high school centile rank alone. These correlation coefficients were reported significant at or above the .01 level. Ernest was able to predict college success without using any preservice music test. Therefore, the single best predictor, even for music students at the college level, was and is today the high school centile rank (Ernest 1970:273-276).

Whellams (1970) administered the Wing Tests of Musical Intelligence and used the subtest scores together with recruitment tests data: picture intelligence, math, spelling, verbal ability, series, mechanical ability, arithmetic, algebra, spatial (squares), length of previous experience, terms in school, age in months, time when first studied piano, time when first studied stringed instrument, and brass instrument, to predict success for both junior level musicians and buglers at the Royal Marines School of Music. Standards at entrance were lower for buglers necessitating the two experimental groups. While the chord analysis, rhythm, intensity, and phrasing subtests of the Wing battery were insignificant, useful predictions could be formulated using the Memory, Pitch Change, and Harmony subtests for success in instrumental music.

> Whilst the results suggests some justification for the addition of non-musical tests to aptitude batteries, and particularly to Wing's, there exists a doubt as to which would be most suitable. Considering the results of both experiments together, it seems likely that the most suitable non-musical tests would probably vary according to the social and educational background of testees, and according to the proportion of the total distribution intended to be selected (Whellams 1970:21).

Whellams was relatively thorough and used all available experimental variable data. The criterion variables included grades awarded by the teaching staff over a period of ten months which identified below average and above average students. As will be mentioned later, the discriminant analysis technique which Whellams used is a very good one for group data. However, the way in which he used it, for all practical purposes, really breaks down into a straight multiple regression approach because the weights obtained are similar to regression weights: this only holds true in the two group discriminant function case, however (Tatsuoka 1971:173).

Lee (1967) established normative data for midwest college and university freshman music students on the Gordon Musical Aptitude Profile and also investigated the reliability of the battery for college and university music students and investigated the relationships between freshman music students' test scores and various course grades as well as the relationships between college choir members' test scores and their musical aptitude as rated by their choir director. Lee concluded:

> Resultant data suggest that although the Musical Aptitude Profile was designed for use with public school students, it can serve as a relatively effective musical aptitude test for use with college and university music students. This conclusion rests on the fact that (a) the test scores are reasonably reliable for college and university freshman music students, (b) the test scores display considerable variability for college and university freshman music students, (c) the various subtests and total tests measure quite different dimensions of musical aptitude for college and university freshman music students, (d) test results are more related to musical endeavors than to academic achievement for college and university freshman music students (Lee 1967:287).

With 57 subjects at a university in western Illinois, Lee found the following correlations for MAP subtests and total score to the music theory grade: Tonal Imagery (T) r=.32 at the .05 level, Rhythm Imagery (R) 5=.19 not significant, Musical Sensitivity (S) r=.36 at the .01 level, and Total Test Score (MAP) r=.35 at the .01 level. Lee concluded that significant relationships of the MAP battery were useful at the college level with freshman music majors.

To test this hypothesis as well as to establish validity for a college music entrance test battery, the investigator (Krueger 1972) used the MAP battery and the Aliferis Music Achievement Test - College Entrance Level with 42 freshman music majors at Oklahoma City University. The tests were given during the first two weeks of the 1972-73 school year. A ranked theory grade was offered for the criterion variable by rank ordering all three theory teachers' students within each class from high to low and dividing each class into sixths. Thus, a sixway scale was transformed for the ranking with 14 students in each category from one through six. This forced a better distribution of scores than would have been achieved by using the theory class grade average for the freshman year since these tended to be negatively skewed. With this ranked criterion, the following correlations were obtained:

Music Test	Theory Ranked Grade
Aliferis 1	.53**
Aliferis 2	.04
Aliferis 3	.31*
Aliferis Total	.22
MAP (T)	.16
MAP (R)	-.14
MAP (S)	-.12
MAP Total	-.19

p<.01**, p<.05*

The relationship between MAP Total and Aliferis Total was r=-.26 significant at the .05 level. The Sixteen Personality Factor Questionnaire and the Motivation Analysis Test were also used in this study. Factor B

(Bright versus Dull) correlated .27 at the .05 level for the MAP Test and -.04 to the Aliferis Total score. The purpose of the study was to determine the usefulness of the MAP and Aliferis Tests for advising and counseling freshman music majors at Oklahoma City University. The general conclusions was that the Musical Aptitude Profile test battery did not discriminate sufficiently among the students to warrant its lengthy inclusion in an entrance testing battery. The Aliferis test appeared to be more discriminating (Krueger 1973:9).

Lee's study has value for further replication while the investigator's study might have been biased due to a highly select group of entering freshman music majors. As with any entering testing program, the battery should be validated yearly because of possible student and university changes.

In a pilot study to this research project, the investigator (Krueger 1971) studied the relationships among musical activities, non-musical activities, and work activities and the cumulative grade point averages of high potential, average potential, and low potential undergraduate music education students at the Urbana campus of the University of Illinois. A second purpose of the study was to develop regression equations useful for predicting cumulative grade point averages from these measures of activities.

The undergraduate music education students were placed into three experimental groups on the basis of their high school centile rank: (1) high potential - 95 to 99 centile ranks, N=27, (2) average potential - 74 through 94 centile ranks, N=46, and (3) low potential - 22 through 73 centile ranks, N=27, with a total group, N=100, of randomly selected sophomore, junior, and senior music education students. No differences for sex or major were investigated in this pilot study.

The data gathering instrument was a one page checklist which recorded the following information from the students' files located in the School of Music's Department of Music Education office: student number, sex, major, year in school, high school centile rank, cumulative grade point average, hours in musical activities per week, hours in non-musical activities per week, and hours in work and employment activities per week. For each of the three artificially dichotomized variables (musical, non-musical, and work activities), an unweighted means analysis of variance was employed in a three by two factorial design with high school centile rank (high, average, and low potential). The criterion or dependent variable in each case was the cumulative grade point average. A stepwise multiple regression analysis using three independent variables on the cumulative grade point averages was also undertaken for each group.

There were significant differences in cumulative grade point averages for high potential ($p \geq .001$), average potential ($p \geq .001$), and low potential ($p \geq .002$) groups of music education students. Second, there were significant cumulative grade point average differences between high, average, and low potential groups who were active or inactive in musical activities ($p \geq .003$). Third, there were no significant differences for the groups on cumulative grade point averages and non-musical and work activities. Furthermore, there were no significant interactions.

By treating the activity variables as continuous variables in a regression equation, a significant multiple correlation was found between activities and grade point average for only the high potential group ($R=.51$). The multiple regression analyses found a significant negative correlation between work activities and grade point average for the high potential group ($r=-.39$). While there were no significant interactions found in the ANOVA analyses, it appears from the regression analyses that high achieving students in the high potential group were not employed, but, those that did work had significantly lower grade point averages.

The conclusion was that the high school centile rank was a useful variable for generating high, average, and low potential groups for college grade point success and was significantly related to achievement in college. The relationship of interest, activity, and work variables affecting grade point averages is open to further research. This pilot investigation supported the conclusions of Ernest and assisted this study by directing the focus to inservice music educators (Krueger 1971:1-7).

Maier (1970) in a study at Fredonia, New York, of music majors at the State University concluded:

> 1. It is possible to predict examination scores for keyboard harmony, sightsinging, dictation, rudiments, and part-writing and analysis; in addition to total grade averages and grade averages in music theory at the State University College, Fredonia, New York.
>
> 2. The same set of independent variables cannot be used to obtain the optimum prediction of the level of achievement in all criteria.
>
> 3. The high school Grade Average is the most useful independent variable for predicting success in a large number of criteria.
>
> 4. With the exception of the High School Grade Average, the "non-musical" independent variables are not useful for predicting success in the music theory criteria. However, the scores for the Regents Scholarship and Qualification Test are useful for predicting the total grade average criteria.
>
> 5. The most useful of the "musical" independent variables for predicting success in the music theory criteria are the Dictation and Part-Writing scores from the Music Theory Placement Test and the Interval Aural Recognition, Scale Relations Total -- Accuracy, and Chord Notation Total -- Output scores from the Ohio State Diagnostic Music Test Battery: Elements of Music Theory. The most useful "musical" independent variables for predicting the Total Grade Average criteria are Melodic Memory and Seashore Pitch (Maier 1970:99-100).

Maier investigated a number of experimental variables to success in music theory using multiple regression techniques and found that success within the music education curriculum can be accurately and validly predicted from data available at entrance to college. It is interesting how the high school centile rank continues to express itself as an important predictor throughout the literature. This probably best explains the fact that previous success is the best predictor of potential success.

Probably the best longitudinal prediction program in the nation at the college level is currently underway by Poland at The Ohio State University. Since 1956, freshman class applicants have been required to complete the Ohio State Music Test Battery, nearly 1,000 students through 1963. These students were also required to meet all other university requirements for admission.

Poland developed the Ohio State Diagnostic Music Test Battery in 1952 which was based on the limited experimental literature concerned with the analysis of elemental musical behaviors as well as the kinds of tasks the Ohio State School of Music faculty thought students ought to be able to do at the time of admission as freshman music majors. Poland made no attempt to distinguish between the inborn and the learned in his behavior samples. "It is a fact, however, that so far we have discovered no category of musical behavior which is not modifiable by training" (Poland 1963:1).

The test is divided into three test sections. Test One, Elements of Music Theory, is designed to measure certain responses to the elemental categories of intervals, scale, and chords presented in aural and visual form. The visual section uses tachistoscopic presentation. He feels that the visual aspects of musical behavior may be as important as the more commonly accepted aural and notational skills (Poland 1963:2).

Section Two, Music Recognition Test, measures the student's recognition of a musical composition by asking him to select both the composer and the title from what he has heard. Section Three, General Music Information Test, measures knowledge of musical names, places, dates, and musical terms.

The three tests of the battery and the Ohio State Psychological Examination as well as other data from cumulative records were incorporated into a prediction program designed to predict success in the first course in music theory and to place students in the regular or remedial situation. Poland also began to predict success in the School of Music undergraduate degree programs in terms of sixth quarter and twelfth quarter total accumulated grade point hour ratios. In 1962, a satisfactory sixth quarter prediction became a prerequisite for admission to the School of Music. Furthermore, freshman scholarship applicants to the School of Music were required to meet the basic University entrance requirements, take the test battery, and have better than a 50 percent probability of graduating with a B average in order to obtain a scholarship. Of those in the recommended group, over two-thirds kept their scholarships while of those in the non-recommended group, less than one-third kept their scholarships.

Poland found that those applicants with only a vocal background did significantly poorer on all parts of the entrance battery than did those vocal students with some instrumental background. This indicated that the probability for success in the University was significantly lower for vocal music majors (Poland 1963:2-3). The multiple correlation coefficient that Poland derived originally between the test battery-entrance variables and first quarter theory grade was R=.83 (N=91). In a correlation with the predicted and first quarter music theory grades with the actual grades for freshman music majors, a kind of cross validation procedure, Poland obtained an excellent relationship which was highly significant r=.65 (N=37).

To the present day, Poland continues to modify the regression equations used to predict the theory quarter grades and continues to cross validate his findings with additional samples of entering college freshman music majors. While today, the prediction of a B average is no longer a University requirement, probably due to administrator bias and discrimination, Poland has continued to show that a well validated test battery and a sophisticated prediction design can, to a very reliable and useful level, predict entering freshman music majors' theory grades as well as 6th and 12th quarter grade point averages. Poland's research has been the longest and best of any college's School of

Music and indicates that a good testing program is very useful for advising and counseling as well as for individualizing instruction. Most research in music education has not followed the lead set by Poland much to the detriment of the profession.

Dvorak (1973 in progress), using the Colwell Music Achievement Tests I-IV with freshmen through senior undergraduate students at the University of Illinois and other Illinois institutions has further substantiated the claim that accurate predictions can be made for music majors with course grades and grade point averages as criterion variables. Very significant multiple regression correlations have been formulated indicating not only the power of the predictions but also the usefulness of the MAT Battery which was originally designed for elementary and secondary school pupils.

This section has highlighted research projects undertaken to predict success for music majors in the undergraduate music curriculum. For a more complete discussion of the related research in this area from 1929 to 1970, the reader is referred to the Maier (1970) study. His synopsis of the research during this period follows:

> 1. Scores on personality and vocational inventories generally were not related to success in music theory . . .
>
> 2. Findings concerning the high school grade average were not conclusive . . . (Research today has been more conclusive.)
>
> 3. Scores on academic aptitude tests generally were related to success in music theory . . . However, musical tests usually were found to be more significantly related to success in music theory . . .
>
> 4. Of the older, standardized music tests, the several forms of the Seashore Measures of Musical Talents were most used. The relationship of scores on the Seashore tests to success in music theory varied, depending on the criterion measure used. Scores on the battery generally were more related to achievement in the areas of sightsinging and dictation . . . than to achievement in general theory or harmony work . . .

> 5. Other, older, standardized music tests used in the reported studies were the *Tonal Memory Test* from the *Drake Musical Aptitude Tests*, the *Kwalwasser-Dykema Music Tests*, and the *Kwalwasser-Ruch Test of Musical Accomplishment*. Of the *Kwalwasser-Dykema Music Tests*, the sections which related most to success in music theory were the tests of Tonal Memory and Pitch Imagery. The entire battery, however, was not as effective as the Seashore Tests . . . The Drake test was slightly related to success in college music theory . . . whereas, the Kwalwasser-Ruch Test was not related to success in music theory . . .
>
> 6. Of the newer standardized music examinations, the Aliferis *Music Achievement Test: College Entrance Level*, the Wing *Standardized Tests of Musical Intelligence*, and the Gordon *Musical Aptitude Profile* were the only ones used in the reported investigations. Scores on the first two of these tests proved to be moderately related to success in college music theory . . . while scores on the latter were only slightly related to success in college music theory . . .
>
> 7. Scores on original musical examinations or background inventories were often more effective predictors than were scores in available standardized examinations . . .
>
> 8. The strongest predictions were achieved by finding a small number of effective variables and combining them as a battery by using multiple linear regression analysis . . .
>
> 9. Though the findings have been all-conclusive, they indicate that by combining relevant background information with scores on musical and non-musical aptitude and/or achievement tests, a student planning to enroll in a theory course may be given advisement and counseling as to his potential success. Moreover, the diversity of the findings indicate that a prediction program which is optimally effective for one institution may not be as effective for another (Maier 1970:93-95).

Relevant to this research investigation was the fact that success in college, specifically in the music undergraduate curriculum, could be predicted to a reasonable degree of confidence by using information available at entrance in a multiple regression design. Therefore, it was necessary to attempt to determine whether any of these variables would be useful in eventually predicting music teaching success. Obviously, one must graduate from a college before he can teach music in the schools so factors affecting graduation might have a bearing on music teaching success.

Predicting Student Teaching Success

It follows that if success in the undergraduate music curriculum can be predicted to some reliable and valid degree of usefulness that some of these factors may be incorporated in a design to predict student teaching success in music. While the criterion for success in the curriculum has been some type of grade, two types of criteria appear to be useful for student teaching success: student teaching grade and/or student teaching rating, many rating criteria being a composite of the cooperating teacher's and supervisor's ratings. A number of studies have concentrated in this area, the best of which are discussed below. Because reference to personality variables will be made, these studies will also be discussed later in this chapter in Section Three to the degree that they affected the design of this investigation.

A classic doctoral investigation was conducted by Hazel Nohavac Morgan (1946) who attempted to predict student teaching success for senior music students at the University of Minnesota. She developed statistical profiles for use in counseling and advising undergraduate music education students.

Morgan investigated the relationships among general achievement, musical achievement, musical aptitude, family, personality, and attitudinal factors to potential music student teaching success. She used two procedures for determining success: (1) student teaching accumulative ratings by cooperating teachers, and (2) rankings by University of Minnesota faculty members.

She compiled a profile on student teacher ratings by the 23 cooperating teachers. Because these ratings clustered at the upper end of the scale for each trait rated, she applied the hyberbolic tangent transformation technique. The cooperating teachers differed significantly in the ratings they assigned to any one student teacher. She found the variation of the means of the cooperating teacher ratings from the mean for all ratings larger than the residual variation.

The variation remaining after that assignable to trait differences and teacher differences had been removed became the student teaching rating for each music student teacher. In other words, Morgan used part and partial regression techniques to arrive at a residual rating criterion -- the only one found in the music education literature. She simply refused to take the ratings at face value, sum them, and find a composite mean rating score for each student teacher as have most studies using ratings as a criterion.

Four faculty members who had had the music student teachers in class divided them into four groups of potential music teaching success: (1) outstanding, (2) above average, (3) below average, and (4) doubtful. She used the method of paired comparisons to check the reliability of these estimates. This technique is similar for forcing four responses makes the scale a forced choice scale, allowing no middle ground or average category. The faculty members had to select one of the four categories by rank ordering their students and dividing them into fourths.

Morgan investigated 22 traits and abilities, the experimental variables. Each one was described as to range, quartile, and median. Published instruments included the American Council on Education Psychological Examination for College Freshman, the Seashore Music Aptitude Tests, the Minnesota Personality Inventory (MPI, the forerunner to the MMPI), and the Cooperative English Examination. She developed the Music Adjustment Inventory (MAI) to measure attitudes.

Morgan established validity for the MAI by (1) experts' judgments, and (2) the MAI's ability to discriminate between two student groups known to be different in musical participation. Reliability was established by using two forms of the test (written and oral) in a test-retest period of two weeks which provided reliability of .69. Her sample included 122 senior and 40 junior music education majors. Of the following 22 traits and abilities listed below, the first 14 were found to be significant at or beyond the .05 level and related to music student teaching success:

1. High school centile rank
2. American Council Psychological Examination Score
3. Cooperative English Examination
4. MPI - Morale
5. MPI - Social Adjustment
6. MPI - Family Relations
7. MPI - Emotionality
8. MPI - Economic conservation
9. Seashore - Pitch
10. Seashore - Loudness
11. Seashore - Time
12. Seashore - Timbre
13. Seashore - Rhythm
14. Seashore - Tonal Memory
15. *Music Adjustment Inventory*
16. Music education GPA
17. Junior education sequence GPA
18. Music courses GPA
19. Applied music courses GPA
20. Academic minor GPA
21. Student teaching GPA
22. Total University of Minnesota GPA

Morgan found that the ranking procedure to be a more valid and accurate method for predicting student teaching success than ratings. While this ranking method may have succeeded at the University of Minnesota in the 40's, a university with a larger number of students in music education might find it difficult to implement.

She found that students with above median scores on the aforementioned 14 independent experimental variables during their freshman year tended to maintain that superiority throughout the music education curriculum while students with below median profiles tended to stay below the median throughout the music education program.

Morgan's study is superb from the standpoint of thoroughness, comprehensiveness, and scholarship. All her correlations and non-parametric significance tests were computed by hand. One finds few studies today of such magnitude. It is interesting to note that no recent study in this area has included Morgan's four year longitudinal project in the related literature chapter. Perhaps this thesis was too difficult to obtain; yet, Cady in the 1967 issue of the Journal of Research in Music Education discusses the Music Adjustment Inventory that Morgan developed.

Duda (1961) studied the teacher education curriculum with the objective for predicting student teaching success for University of Illinois music education students. Success was defined to be those qualities and characteristics measured by the standardized Ryans Classroom Observation Record. Three major dimensions of teacher classroom behavior are assessed: (X) Pattern X includes the friendly, understanding versus aloof teacher. (Y) Pattern Y refers to the responsible, businesslike, systematic versus the unplanned, evading and slipshod teacher, and (Z, Pattern Z refers to stimulating, imaginative, surgent, versus the dull routine teacher. He found that he could predict to some degree of accuracy Patterns X and Z.

The Classroom Observation Record consists of 18 bipolar dimensions of teacher behavior and 4 of pupil behavior. The rater selects between these bipolar words using a 7 point scale. While Ryans developed this test for use in regular classrooms, it has not been validly researched in music classes. Duda's and Morgan's studies will be discussed in more detail in Section Three.

Duda used the MMPI as one of his predictive instruments and was able to show some predictive relationship between personality variables and the behaviors measured by the Record. Using multiple regression techniques, he found significant multiple R's.

Turrentine (1962) used The Ohio State University Psychological Test, College Entrance Examination Board's Scholastic Aptitude Tests (SAT) - Verbal and Mathematic, Otis' Quick Scoring Mental Ability Test: Grades 9-16, high school graduating class centile rank, grade point average of college sophomore year, GPA of 16 hours of required music theory, and GPA of teacher training courses completed for enrollment in student teaching to predict the student teachers' numerical grades earned in the Conservatory of Music at Lawrence University. Using multiple regression analyses, he found that the best single predictor of success in student teaching at Lawrence University was the grade point average of the teacher training methods courses. He also found that the best two variable combination was high school centile rank and teacher training methods courses GPA. It is interesting to note that none of the other variables were significant nor contributed to the multiple R.

Durflinger (1963) attempted to (1) assess the validity of certain instruments designed to measure the personality of college students as predictors of effectiveness in student teaching and (2) determine which personality traits were most highly correlated with specific characteristics of teaching effectiveness. For his variables, he used: (1) a teacher rating scale, (2) grades in student teaching, (3) grades in methods courses, and (4) total GPA averages as well as parts and combinations of the criteria.

Data gathering instruments included the *California Psychological Inventory*, *Heston Personal Adjustment Inventory*, *Minnesota Teacher Attitude Inventory*, the Elementary Teacher and Masculine-Feminine Scales of the Strong *Vocational Interest Blank for Women*, the *American Council on Education Psychological Examination*, the *How I Teach* by Hadley and Perkins which is an anlysis of teaching practices, and the Personal-Social Scales of the Roeder *Occupational Aptitude Test*. The sample included 130 women and 20 men elementary education majors who were student teaching.

> The successful teacher is not different from members of the standardization sample in dominance, social initiative, and capacity or desire for social status. He tends to display to a significant degree an outgoing, sociable, and participative attitude (Durflinger 1963:389).

Using the four highest correlations with each criterion, Durflinger found a multiple R of .37 with the rating scale and .67 with student teaching grades. However, he cautions and rightly so, that he did not cross validate (check out his equations on another sample for further validity verification) and that this is required for selection purposes. The significance of this study, while outside the field of music education, is that he was able to find a significant relationship of personality to student teaching success. It is also interesting to note that had he included all his experimental variables in the multiple regression analyses, he might have found different relationships due to suppressor variables which might not show up in the original correlational matrix. In any case, personality was significantly related to student teaching success.

Anderson (1968) using the 16PF to determine the predictive value of selected factors to student teacher success at Western Carolina University. He concluded that the best individual predictor was the total grade point average, similar to Turrentine's finding of methods course grade averages. He concluded, "A study of the predictive factors appearing in the equations for the various groups warranted the conclusion that personality factors played an important role in the success of these student teachers" (Anderson 1968). Yet, while playing an important role, he was unable to further differentiate them.

Anderson (1965) had originally desired to predict success for student teachers in music at the University of Southern California. He investigated the relationships of musical talent, personality, and vocational interest factors to music student teaching success. The purpose of his study was to find existing evaluative instruments which measure qualities that leading American educators have considered important characteristics of successful music teachers (experts' opinions). Further, he determined the relation of these various quality measures to successful student teaching in music, and finally, he determined which of these characteristics had value for purposes of prediction.

The criterion for music student teaching success was the composite rating of the cooperating teachers' and the USC supervising teachers' for each music student teacher. He conducted a pilot study to determine the characteristics of successful music teachers and another one to determine the characteristics of tenure music teachers. These teachers did not complete all the tests nor did he compare their scores to unsuccessful music teachers.

The data gathering instruments included: (1) the Guilford-Zimmerman Temperament Survey (GZTS), (2) the Kwalwasser Music Talent Test, (3) the Project Potential Tests of Creativity, and (4) the Guilford-Zimmerman Interest Inventory (GZII). The following experimental variables were investigated:

1. Musical Talent
2. General activity
3. Restraint
4. Ascendance
5. Sociability
6. Emotional stability
7. Objectivity
8. Friendliness
9. Thoughtfulness
10. Masculinity
11. Personal values
12. Creativity
13. Interest

His sample included 10 high potential and 9 low potential music student teachers as based on the composite rating. Nonparametric statistics included the Chi Square and Fisher Exact Probability Procedure techniques for significance testing of the null hypotheses. Median scores were computed for each experimental variable. No regression analyses were attempted because Anderson felt the sample to be too small.

Three factors in the battery showed possible predictive value since they revealed a difference between the high potential and low potential music student teacher groups: (1) the objectivity factor GZTS at the .05 level, (2) the masculinity factor GZTS at the .05 level, and the (3) symbol production factor of creativity from the Project Potential Tests of Creativity at the .05 level. Music talent from the KMTT was at the .15 level and emotional stability at the .10 level. He concluded:

> The GZTS, the KMTT, and the Symbol Reproduction Tests of Creativity were found to define some significant differences between more successful and less successful student music teachers. Whether or not these relationships are sufficient to assure predictive value in these instruments was not demonstrated and must await further research (Anderson 1965:85).

Wink (1967) investigated the relationships between the self concept, personality need system, and anxiety level of music student teachers and their achievement in the student teaching experience. He also investigated (1) the possible effects of the student teaching experience on music student teachers' self concept, personality need system, and anxiety level; (2) the relationship of changes in self concept, personality need system, and anxiety level to achievement in music student teaching; and (3) the comparison of music student teachers with established norms for self concept, personality need system, and anxiety level. The subjects included 32 music student teachers at The Ohio State University School of Music during the 1966-67 school year.

He used the Bills Index of Adjustment and Values (IAV), Edwards Personal Preference Schedule (EPPS), and the IPAT Anxiety Level (IPAT). In addition he constructed a measure of the phenomenological aspects of the music student teaching experience: the Self Concept of Music Teaching Ability (SCMTA). The Stanford Teacher Competence Appraisal Guide was selected as the criterion instrument and provided a score for the subjects' beginning and final music student teaching successes. A team of three college supervisors and the cooperating teachers observed the student teachers both at the beginning and during the final week of the student teaching experience. The difference between the first and last observations was used as the degree of growth.

While this is not a prediction study per se, Wink attempted to find significant correlations among the above variables and success groups of music student teachers. He used both parametric and nonparametric tests: the Pearson r, t-tests, Spearman rank correlation coefficient, Wilcoxen match pairs signed ranks test, and Chi square. He found the following significant at or beyond the .05 level:

1. There was a positive significant relationship between self concept of music teaching ability and achievement in student teaching.

> 2. Significant changes occurred during the student teaching experience for the LAV and the SCMTA, but not for the EPPS nor the IPAT.
>
> 3. Changes occurring during the student teaching experience for the LAV, SCMTA, and EPPS were not significantly related to achievement; but those who achieved higher in music student teaching indicated significantly less anxiety than those who achieved lower.
>
> 4. Music student teachers with a high self-concept were dependent, conforming, and gregarious.
>
> 5. There were high correlations between anxiety and adjustment - the more well-adjusted subjects indicated less anxiety.
>
> 6. Subjects in the experimental group were significantly higher than the normal population in self-concept, deference, and order; they were significantly lower in affliation, intraception, and anxiety.
>
> 7. High achievers in music student teaching indicated a high need for deference, order and affliation; they indicated a low need for autonomy. There was no difference between high and low achievers for anxiety level (Wink 1970:240-241).

This study will be discussed in more detail in Section Three of this Chapter but it does show the differences for certain personality and self concept variables between successful and less successful groups of music student teachers. It indicates that predictions could be made before student teaching which may be fairly accurate and reliable as to potential music student teaching success.

Smith (1969) like Wink discovered that student teachers' confidence levels toward music and music teaching increased after the student teaching experience. An interesting finding was that a significant negative change in attitudes toward children and teaching occurred during the student teaching experience as measured by the Minnesota Teacher Attitude Inventory (Smith 1969:155).

If one could determine the most common difficulties encountered during student teaching and music teaching activities, it may be possible to devise a prediction program which looks for variables that may be related to those difficulties. Esselstrom (1968) devised a study to determine exactly what

BEST COPY AVAILABLE

college supervisors of student teachers in music can, should, and actually accomplish. While his emphasis was on the role of the supervisor, he found the following opinions about the music student teachers:

> The persistent or common difficulties encountered while supervising student teachers in music include musical problems related to conducting, ear training, and instrumental skills; lack of classroom discipline or communication with the children; the narrowness of music student teachers; and the scarcity of qualified cooperating teachers. The major techniques used to solve these problems are conferences, seminars, careful selection of cooperating teachers, observation, lesson plans, self-evaluation, and demonstration lessons (Crews 1970:37).

It is interesting to note that listed first are difficulties related to skills taught in music theory classes and conducting classes. Perhaps a program for predicting student teaching success might look at the theory variables and incorporate these in regression equations.

An earlier study which attempted to predict potential cooperating teacher success was completed by Williams (1958) at the University of Illinois. He determined the relationships of cooperation teacher effectiveness criteria to selected factors of musicianship, attitude, and personality.

He derived multiple regression equations for the prediction of cooperating teacher success. This study was unique in approach as it attempted to refine a very crucial area of teacher training in music. It is likely that the cooperating teacher has the single most influence over how a music student teacher will act in the field. Williams developed a Student Teaching Inventory which was mailed to cooperating teachers' former music student teachers. Positive correlations were found between these opinions of cooperating teachers' effectiveness with ratings given them by the University of Illinois supervising teachers.

He used the MMPI, the IPAT Music Preference Test which measure personality traits via reactions to musical selections, and college records to gather data for his investigation. While he found some significant relationships between success and psychological factors, practical limitations dictate that less able cooperating teachers may have to be used because of travel and availability considerations. Esselstrom emphasized this point about cooperating teachers' effectiveness, too. Williams' purpose in refining this selection process was admirable and indicates that cooperating teachers' success can be predicted. He was able to derive significant multiple correlations.

This section so far has highlighted the recent research investigations in the area of predicting music student teaching success. Many of these studies will be discussed in more detail in Section Three of this chapter as they relate to the specifics of the design of this present investigation. In summary, a selected survey of the literature shows that music student teaching success can be predicted to some degree of usefulness and accuracy using multiple regression techniques. The most important variables appear to be college experiences as indicated by grade point averages in music and other courses, high school centile rank, selected personality and motivational variables, and certain measurable musical variables. It is important to reiterate that prediction programs incorporated at one teacher training institution cannot be used at another without further local validation due to the uniqueness of each institution's music majors. Therefore, these findings in the literature suggest the importance of validating predictive regression equations in the local setting.

Predicting Teaching Success

There are numerous studies in the literature relating to predicting teaching success. The first category of study involves survey effectiveness investigations which attempt to ascertain the benefits of a particular college's undergraduate teacher preparation program as far as the opinions of its graduates are concerned. These studies, while not predictive nor evaluative per se, are of value because they solicit the opinions of inservice teachers. The assumption is that, by maintaining themselves in the teaching profession, these teachers have achieved a certain amount of success. Therefore, it is reasonable to sample their opinions and views as these may affect future teachers. The weakness of these studies is that all teachers are considered equal in success; therefore, no stratification of teachers into effectiveness groups can be made for comparisons of group responses to questionnaire or opionnaire items. Few studies have attempted this approach.

A second category of prediction studies with teachers may be classified as relationship or correlational studies. These have investigated certain preteaching variables' relationships to teaching effectiveness criteria. These are discussed in this review and are often called prediction studies when, in fact, they are not formulating predictions.

A third area or category might be classified as testing instrument developmental studies. The usual purposes of these studies are to develop testing instruments which are reliable and valid for some purpose of identifying successful teachers.

With prediction studies in teacher effectiveness, everything hinges on the criterion selected. The criterion domain is discussed in detail in Section Two of this chapter. Generally, studies have used one or a combination of the following criteria: (1) pupil gain scores, (2) ratings of teaching effectiveness, (3) ratings related to classroom environments and interactions as measured by the Ryans _Teacher Classification Record_ and the Flanders _Interaction Analysis Scale_, and (4) various process kinds of criterion variables (descriptors of what occurs or doesn't occur). It can be stated to some degree of confidence that teaching success can be predicted.

Barr, Torgerson, and others (1935) used three tests of personal characteristics and one personality rating scale based upon 33 personality traits and found that none of the measures, except possible social intelligence with an overall composite criterion, and the personality rating, was statistically significant. The latter correlation was probably due to the halo effect and the former to some sort of generalized verbal aptitude. They concluded that the three criteria employed appeared to measure different things and were not useful in predicting teaching success.

Erickson (1954) employed seven measures of personal characteristics derived from the Thurstone Temperament Schedule and the Sixteen Personality Factor Questionnaire and correlated these to supervisory ratings, self ratings, peer ratings, and pupil ratings. Consistently low coefficients of correlation were reported and did not differentiate or predict success.

Singer (1954) investigated the relationship of social competency to teacher effectiveness as defined by a supervisory rating criteria. Statistically significant and positive correlations, while not high, were found which offer some promise for further investigation.

Schwartz (1950) developed from the 16PF new tests to cover jokes, tempo, absence of questionable preferences in readings, perceptual speed of closure, inability to state logical assumptions, suggestibility, cube perspective, ideomotor speed, ratio consonant/dissonant statements recalled, ratio purposeful to chance observation and memory, two hand coordination, immaturity of opinion, impairment of memory by emotion, reaction time, body sway, suggestibility and perseveration with supervisory rating as the criterion. A multiple correlation (R=.74) was found for three measures: two hand coordination, reaction time, and the Washburne Social Adjustment Inventory and the rating criterion. Montrose (1954) followed by Schwartz's study and found substantial agreement with his findings which makes them somewhat more interesting from a predictive standpoint than the results from isolated studies.

Jones (1956) correlated scores on a number of measures of temperament to a composite rating. The scores on measures of disposition rigidity, flexibility, general activity, restraint, ascendance, and sociability correlated significantly with the criterion. Barr states, "The findings are significant because they . . . indicate that temperament measurement may be a fruitful field for predictors of teacher effectiveness" (Barr 1961:105).

Flanagan (1961) investigated the Minnesota Multiphasic Personality Inventory (MMPI) in relation to teaching effectiveness and found hysteria and masculinity significantly related to teaching success. He also found an inverse relationship for depression for women. Looking at the patterns of responses, he found that men scored significantly different from women. He concluded that the MMPI has good possibilities as a predictor of teacher effectiveness and deserves wide use as a screening device and a clearance test. It should be pointed out that the MMPI is a pathologically oriented examination which attempts to separate normals from abnormals. It has been used extensively in the literature thus supporting some of Flanagan's predictions about its potential.

Shanks (1969) in a study of concept achievement in science and patterns of teaching behaviors found that high achievers demonstrated certain personality traits which set them apart from low achievers. These high achievers were goal oriented women and self directed while low achievers lacked self discipline and self confidence. "These findings indicate that more attention should be given to and more of an issue made for the identification and evaluation of the personality functions related to successful (science) teaching behavior" (Shanks 1969:28).

Lamke (1951) concluded that the responses of good and poor teachers did not fall into two well defined and characteristic patterns as measured by the 16PF. Several patterns exist for good teachers and probably several more for poor teachers.

Because criterion development has been the major obstacle to predicting teaching success, many studies have concentrated in this area. These are discussed in Section Two. Generally, it can be concluded from studies in the period 1924 to 1961 that certain aspects of teaching success can be predicted to some degree of usefulness. The crucial point is the selection of the criterion variable. Even so, factors which appear to be significantly related to teaching success are: college success in various areas, family background and socioeconomic level, personality and attitudinal factors, and intelligence. For a more thorough discussion of the literature in the general education area of predicting teaching success, the reader is referred to the writing of A. S. Barr (1961).

Predicting Music Teaching Success

Benner (1962) investigated the relationships between (1) an inservice rating of music teachers (the criterion variable), and (2) the teachers' grade point averages in ten selected undergraduate areas as well as their centile rank on The Ohio State Psychological Test (the predictive variables). The sample included 100 graduates of The Ohio State University who had received a bachelor of science degree in music education between 1958 and 1961. At the time of the study, these teachers had taught from 1 to 4 years.

He found the following significantly related to music teaching success: (1) student teaching grade point average ($p \geq .01$), (2) music education methods grade point average ($p \geq .05$), and professional education sequence GPA ($p \geq .01$). None of the other 8 variables was significant.

A similar study was conducted by Borkowski (1967) who found little relationship between the quality of work in undergraduate courses leading to a music education degree and teaching success. He investigated 53 West Virginia high school instrumental music teachers who had at least 3 years of teaching experience. His criterion variables included: (1) pupil performance in which 8 pupils per teacher were tape recorded and the performances judged by a jury of experts, (2) pupil knowledge in music history and theory, and (3) band performance by a jury of experts who heard festival tapes of the directors' bands. He also used a self rating as an experimental variable: the band directors rated themselves as to their playing proficiency during college on their major instrument.

Borkowski concluded that lack of interest in pupils or in teaching may result in less effective teaching while quality of undergraduate work is in no way a measure of the teacher's adaptability or flexibility to adjust to different needs, persons, and situations. "Attitudes, rigid value systems, and personality are all factors which may influence the effectiveness of the teacher more than the quality of his work in undergraduate courses" (Borkowski 1970:143-144).

Strub (1957) investigated 105 successful and unsuccessful public school music teachers in Missouri and found that superior teachers excelled in college grades in general as undergraduates as well as in age and experience, years of marriage, number of college degrees, and musical training before college.

Lutz (1963) investigated the professional backgrounds, experiences, and personality characteristics of 103 instrumental music teachers in Kentucky. He developed an Opinion Rating Form for administrators, fellow teachers, and students to use in evaluating the instrumental music teachers and an Instrumental Music Teacher Questionnaire to gather data about the teachers. Participants furnished the names and addresses of their administrators, fellow teachers, and students. One rating was gathered in each category and a composite rating was used for the criterion variable. The MMPI gathered personality information useful to his study.

The 103 high school instrumental music teachers were stratified into two groups: (1) successful (N=75) and (2) unsuccessful (N=28). Mean profiles for the groups were established on variables of professional background, experiences, and personality characteristics. The Chi Square technique was used to test for significant differences between the groups on each experimental variable.

Lutz found that successful high school instrumental music teachers tend to have less college teaching experience than the unsuccessful group. The successful group experienced less solo competition at the high school level and was somewhat younger and more satisfied than the unsuccessful group. He found significant differences in personality characteristics in three MMPI areas: (1) depression, (2) psychopathic deviation, and (3) psychasthenia. From this analysis, he concluded that successful high school instrumental music teachers: (1) are less committed to compulsive or neurotic behaviors, (2) are less moody and better able to reconcile their internal problems and concerns, (3) are more capable of deep emotional response, (4) are better able to profit from experience, (5) have a higher degree of emotional morale and are better able to manifest normal optimism with regard to the future, (6) strive to a higher degree for social approval, (7) are less hostile, (8) worry less and are more flexible in meeting day to day situations, (9) have broader interests, (10) are less subject to introversion and tend to be more people oriented, (11) tend to be happier, (12) are more self satisfied with regard to their professional and personal lives, and (13) are more self confident than unsuccessful instrumental music teachers (Lutz 1963:117-118).

These differences, while significant, were in degree rather than in direction. The MMPI group profile for the successful group was very similar to the unsuccessful group. The unsuccessful group scored higher on 14 scales which indicated that the successful group had a lesser degree of undesirable personality characteristics (Lutz 1963:121).

Bullock (1973 in progress) is investigating selecting personal and professional qualities common to a group of superior junior high school instrumental music teachers in New York state excluding New York City. This is a further attempt to refine Lutz's findings. The basic assumption being tested is that successful music teachers have a pattern of identifiable attributes which, when measured, will reveal a model profile of a successful junior high school instrumental music teacher. The strength of such attributes as personality, attitude, self-concept, and work satisfaction as well as the interactions between these traits should reveal this profile which he feels is undefined at the present time (Bullock 1972:1).

Bullock is using an interesting criterion - peer nominations - for determining his successful group of instrumental music educators. The sample will be nominated by high school instrumental music teachers in New York state and by experts. A total of 53 counties and 102 high school instrumental music teachers, and two experts will be involved in the selection of the junior high school music teacher sample.

Bullock is using four measures to gather data for the experimental variables: The _Training and Experience Questionnaire_ (TEQ) developed by him to examine personal and professional training and experience, the _Personal Interview Questionnaire_ (PIQ) also developed by him to examine in depth qualities of training, experience, personality, and job satisfaction, the _Sixteen Personality Factor Questionnaire (16PF) Form A_ to examine basic personality traits, and the _Minnesota Satisfaction Questionnaire_ to examine employment satisfaction. He is tape recording the interviews for further analyses and plans to use multivariate analysis of variance techniques. His study appears to be a well planned and a much welcomed investigation.

Fosse (1965) investigated criteria which could be used to predict the future teaching effectiveness of persons intending to become high school band directors. What he actually did was to develop profiles for four categories of band directors. He investigated the relationships among five categories of affective variables to the four groups of band directors. These were:

1. Biographical characteristics of the subjects including family background and marital history.
2. The subjects' musical education and experience before, during, and after college.
3. The subjects' teaching experience.
4. The subjects' psychological attributes at the time of the study.
5. The factors in the subjects' teaching environments at the time of the study (Fosse 1965:2).

The success criterion which was used to stratify his sample was the rating awarded to each band director in the 1961 Illinois State Music Festival. He also investigated band directors who did not enter their bands in the Festival.

The sample included volunteers teaching in Northern Illinois for at least 3 years: (1) Set I was composed of seven band directors who received a superior rating, (2) Set II were those receiving an excellent rating (N=10), (3) Set III band directors received a good rating (N=5), and (4) Set IV directors did not enter the festival (N=8) for a total sample of 30 high school band directors. None of the subjects was aware that he was being assigned a rating of music teaching success relative to the Illinois Festival rating of his high school band. While one may criticize this limited definition of music teaching success, it does make sense. It is probably no worse than using a composite rating from the building principal and music supervisor. These Festival ratings were awarded by three music adjudicators independently using the same scale rating form and then summed for a single rating criterion.

Nonparametric statistics were used in this study: Kruskal Wallis One Way Analysis of Variance, Fisher Exact Probability Test, and the Chi Square Test. One of the staff programmers at Northwestern University suggested the use of the discriminant function for comparing differences among the four groups and to produce weights useful in predictive equations (Fosse 1965:82). This technique had not been used at Northwestern with a small group of subjects until this study.

Out of a total of 350 individual variables gathered in the 5 affective categories, 33 ordinal variables were found to be significant at the .05 level. Two verbal analyses of the data were made and included noninterpretative sketches of the four sets of subjects and the psychological interpretations of coded profiles of the typical subject from each set by a psychologist.

Fosse used the MMPI, college records, and checklists to gather data. He interviewed each subject as well as visiting the schools in which they were teaching music. With few exceptions, Fosse found that most of the variables selected for use in the equations were concerned with the subjects' psychological scores, basic personality traits. The most noticeable exception was the subjects' fathers' occupational status (Fosse 1965:269). There have been many indications from Cattell's research to indicate that significant relationships between home environments and personality variables exist.

Set I band director's basic personality referred to aloofness with well repressed hostility. He used his profession as a source of security, was energetic, ambitious, and methodical in his work. Administrators would look upon the Set I director as an ideal type of teacher. Fosse concluded that the Set I director would appear to be a technician, pleasant and easy to talk to within an aura of restraint. With his students, he will exhibit calm, self control but when seriously frustrated will resort to anger such as yelling, throwing physical objects, and, in some schools, by assaulting students. After this type of behavior, he will rationalize his actions by displacing the guilt onto the students (Fosse 1965:270-272).

The Set II director is more relaxed and warmer than the Set I director. While this may cause him to be better liked, it also may lead him to be less critical of under achievement of pupils. He will be more relaxed in mundane tasks such as ordering, sorting, and distributing music, planning concert programs well ahead of performance dates, and setting up a student officer organization.

The Set III director came from a significantly higher socioeconomic status that sets him apart from the first two directors. He is presently working in an occupation with lower status than his father's. Whether this phenomenon represents an unconscious desire to escape from an authoritarian figure cannot be assumed by Fosse's data; but, psychologically, the Set III director has difficulty in relationships with authority figures. He appears to have difficulty handling his entire situation since he exhibits anxiety and depression. He is considerably disorganized in his teaching, rehearses his band considerably more times and has more sectional rehearsals than the Set I director, receives considerably lower ratings, and in times of higher activity, starts projects or programs that he does not carry through. He is submissive toward anyone with an authority image. "When asked to complete forms, he will do so with alacrity, and may be overly zealous in completing them" (Fosse 1965:274). He will be eager to present a pleasant image on first acquaintance, but his overall dissatisfaction will tend to cause friction inside the relationship.

The Set IV director was significantly older than the others. Psychologically, he does not have the drive of the first two which also substantiates Cattell's findings that age is a significant factor in submissiveness and drive. The Set IV director was the most reluctant to complete forms or psychological tests; and the psychological report noted lack of interest, depression, and uneasiness under supervision or scrutiny (Fosse 1965:275). There was no attempt to define the effectiveness of the Set IV director in relation to the other directors.

Barth (1961) studied the personality and temperamental characteristics of 316 school music teachers, 10 percent of whom were rated best by a panel of experts, using the 16PF and the Thurstone Temperament Schedule. The top 10 percent considered themselves to be most satisfied and differed from the other certificated music teachers in general mental ability, perseverance, persistence, and exceeded the others only in general intelligence. No personality pattern differentation was evident. The total music teacher group was found to be more self sufficient, accepting, outgoing, understanding, permissive, confident, self secure, less tense and anxious, more active, and possessed more leadership qualities and were more emotionally stable than the general population (Barth 1961:107).

Another area of possible importance bearing on the problem of predicting teaching success in music revolves around the area of reasons for which persons remain in or leave the music teaching profession. While it is not assumed that persons leave the profession because of inability, there may be certain motivational, attitudinal, and/or personality characteristics influencing these

BEST COPY AVAILABLE

decisions. If the reasons for leaving or not entering the music education profession are similar, it may be possible for corrective measures to be undertaken during the undergraduate music education experience.

Calder (1962) investigated the causal factors for male music education graduates of 9 Pennsylvania teacher training institutions leaving the music education profession. A secondary purpose was to determine the causal factors for qualified male music education graduates of these institutions not entering the profession. The sample was limited to graduates during the 1950-59 period for questionnaire analyses. A total of 183 respondents completed the questionnaires.

Calder item analyzed his questionnaire and found a general consistency of responses. Furthermore, 15 follow-up interviews provided an additional consistency figure. Through t-tests, he found the following to be most significant:

> 1. Opportunity for other positions offering better salaries, better working conditions, or more prestige.
>
> 2. Opportunities for advancement too limited in public school music teaching (Farwell 1964:49).

Other factors related to equipment, disinterested administration, musical standards, teaching schedule, unconducive student attitude, need for additional employment, lack of administrative respect for the music teachers' abilities, change in standard of living values by the teacher and his family, and dissatisfaction with the speciality of music teaching.

The five most important reasons for graduates not entering the profession were: (1) unsatisfactory maximum salary, (2) low starting salary, (3) opportunity for a position offering a better salary, better working conditions, or more prestige, (4) limited opportunities for advancement in public school music, and (5) musical immaturity of public school students (Farwell 1964:50).

Rhinehart (1963) in a similar study of Florida State University School of Music graduates from 1950 through 1956 received a 72.4 percent return from an original sample of 377 possible respondents. He found that 60 percent of the men and 45 percent of the women teaching music at that time had seriously considered leaving the profession. He concluded that over 70 percent of the respondents felt that inadequate salary, unsatisfactory space facilities, inadequate equipment, poor class scheduling, and an inadequate budget were the prime reasons for leaving the profession (White 1965:45).

Studies by Snapp (1953), Sperry (1958), Michael, Barth, and Kaiser (1961), Bunyard (1963), White (1967), and Bodanske (1969) concluded the same general findings as those mentioned above. If one looks closely at the reasons for leaving the profession, he is impressed by the fact that these teachers should have known about the extenuating and adversive circumstances before entering the profession and before completing the degree in music education. It is quite possible that these teachers simply took the degree in order to assure themselves of employment. Salary, working conditions, scheduling, and other conditions should be brought to the attention of prospective music educators. Recent studies in this area may demonstrate more teacher stability in music due to significantly higher salaries; median salaries doubled in the period 1960 to 1970.

Additional contemplation of the reasons for leaving the profession center round personality and motivational variables. It appears that these teachers: (1) really did not understand what they were getting into as indicated by their reactions to their students, (2) did not appreciate or have a cognizance of the salary levels, (3) did not foresee administrative disinterest and/or conflict, and (4) generally were not prepared for the human relations and living aspects of the music teaching act. From these studies, one can conclude that most of these variables that relate to leaving the music education profession can be determined before entrance, e.g. sometime during the undergraduate curriculum. If these drives, sentiments, and personality traits can be diagnosed early, advising and counseling may be more effective. For this reason, it was decided that personality and motivation would be the primary emphasis of this present research investigation.

A final area of research which may relate to predicting music teaching success centers round the general area of evaluation of undergraduate music education programs. Most of these studies have solicited responses from graduates of the institutions under investigation by questionnaire. These opinions are then supposedly used to revitalize the teacher preparation curriculum. Studies by Wilson (1958) at The Ohio State University and Taylor (1969) of Maryland music educators along with numerous others of lesser quality have one basic fault that cannot be overcome in the design. They did not discriminate between the answers of superior versus less outstanding groups of music educators. Without some kind of criterion, the results of the questionnaire analyses are of dubious value.

A related area of research relating to predicting music teaching success is found in the general area of testing and evaluation. Snapp (1953) using the Strong technique constructed an interest inventory with appropriate scales for use in counseling college level music students. Interest factor items were selected and a forced choice section was developed concerning musical personalities which identified musicians by name in 8 general categories. Also included were items requiring an estimate of job satisfaction for these kinds of musicians. The inventory was validated by criterion groups of college music theory teachers, college applied teachers, public school music teachers, and symphony musicians. Reliabilities reported were .93 for the entire inventory and .82 for the musical personalities section (Cady 1967:141;142).

Watkins (1958) developed a forced choice instrument for use by junior high school pupils to evaluate their music student teachers. Descriptive sketches of favorable and unfavorable traits in music teachers were submitted to a jury of experts for content validity. Today, Watkins has developed and expanded this scale for use in evaluating junior high school music teachers by adult and student raters.

Gray (1960) made initial steps toward a standardized battery of tests in music education as minimal standards examinations. He examined the curricula of 63 state universities and compiled items similar to all. The instrument was validated by faculty opinions. The development of this instrument, according to Cady, may be considered the first attempt toward a standardized music achievement test battery for the accredidation of music teachers on a nationwide basis (Cady 1967:143).

Since 1957, the National Teacher Examination, published by the Educational Testing Service, has included an examination in music education. Lehman emphasizes the fact that these tests have no official status, and, while confidential, are used by a considerable number of school districts. He believes that at least two of the assumptions on which the test is based are to some extent faulty:

> First, it assumes that the music teacher should be more or less equally familiar with all phases of music teaching from the elementary school through the secondary school, instrumental and vocal. Second, it assumes that there is one correct way to accomplish a given task, and that although we must recognize individual differences in students we need not recognize individual differences in teachers (Lehman 1969b:23-24).

The test does not attempt to measure personality, interest in children, teaching effectiveness, musicianship, nor ability to motivate learning. Lehman states that the test may be useful in eliminating obvious misfits through preliminary screening, and it may provide a measure of objectivity when a few teachers are to be chosen from among a large number of applicants. "It is wholly unsuited for use as a criterion for promotion, tenure, or salary increases" (Lehman 1969b:24).

Not to give up easily, the Educational Testing Service in 1972 inaugurated the School Personnel Research and Evaluation Services Program (SPRESP) which includes a "Common Examination" battery comprising professional education, written English expression, social studies, literature, and the fine arts, science, and mathematics. There are also 20 "Speciality Examinations" including one in music education. The program is intended for a variety of purposes such as evaluating the effects of inservice training for educational personnel, and for gathering background data on the educational staff. The exams are similar to the National Teacher Examinations but are more flexible in that they can be given at any time.

The examination in music is designed for the generalist and may be appropriate for the music supervisor but may work to the disadvantage of the outstanding choral or instrumental music teacher. The exam assumes that the teacher will function primarily as a conductor. The members of the Music Educators National Conference Commission on Instruction concluded:

> ETS wisely cautions that these examinations should not be used by themselves and that they measure only a portion of the knowledge and skills required . . . It (the test) suffers from nearly all the shortcomings attributed to objective tests in recent years by critics of standardized testing. A number of test items have no single answer that experts would agree is clearly superior to the others. Other items are based heavily upon oversimplification or generalization. Others are naive; still others are based on trivia. The professional education portion of the Common Examinations is afflicted with many of the ills of the music education examination, though to a much lesser extent. No information concerning reliability or validity is provided for any of the examinations . . . The usefulness of the music education examination, particularly, is extremely limited at best. It surely must be one of ETS's least successful efforts (Lehman and others 1972:413-414).

The Educational Testing Service also publishes a music test as one of the advanced tests comprising the Graduate Record Examination. The content of the test is limited to history and literature which tends to exclude students in composition, performance, music education, and conducting. In concluding this section, no battery of tests has yet been able to discriminate sufficiently among successful and unsuccessful music educators.

> Despite the existence of these tests, all of which are more or less specialized in nature, the need for evaluative devices in music far exceeds the experimental and developmental work being done in the field . . . Music education constitutes a unique blend of skills, knowledge, and affective behaviors, and some of those aspects which are most important also happen to be the

> most difficult to measure. The arts in education are intended to make a contribution to the improvement of the quality of human life, but how can one measure such improvement quantitatively? Traditional measures of income, leisure time, and the number of material conveniences available are certainly quite inappropriate in making a qualitative determination of this nature (Lehman 1969b:24-25).

Section One of this chapter has discussed selected related literature in the areas of predicting college, student teaching, and teaching success. The following are offered as summary conclusions derived from the literature affecting this present research investigation.

1. Success in college can be predicted to a reliable and valid degree of accuracy in criterion areas of specific music courses, grade point averages, ratings, and rankings of success. The best predictor for college achievement is previous achievement usually exemplified by the high school centile rank. Various test batteries of personality, musical aptitude, and musical achievement can further improve the power of the predictions. Underlying these previous behaviors which predict future behaviors are intelligence factors.

Because success in college is a predeterminer of potential success in music teaching, it was decided to use only music educators who had matriculated. In most states, this is a mandatory requirement for certification. Before any research investigation can predict potential success from various preservice variables, it was imperative to investigate successful teachers. One cannot assume that certain achievement factors as exhibited in college relate to music teaching success, the literature being quite inconsistent in this area. The decision was made to narrow the topic of this study to the personality and motivational characteristics of music educators since these seem to affect not only college success but also success in other categories. Intelligence information would be included from one scale (Factor B) of the selected 16PF test. Because of the numbers of music teachers involved in this study, it was determined that it would be very difficult to obtain college transcripts for each participant. Therefore, college experiences per se were not to be under investigation.

2. The literature in the area of predicting student teaching success was varied and inconclusive. Yet, one overriding concern kept evolving. Personality variables when included with other demographic and experiential variables seemed to contribute the most power or account for the most variance in the predictive regression equations and profiles. Studies by Morgan, Anderson (1965), Duda, and Williams emphasized this point while studies by Turrentine and others which did not investigate personality variables and did not produce as many significant findings useful for predicting success.

One of the problems in the literature appeared to be the selection of the criterion variable, usually a composite rating, ranking, or grade in student teaching. When grades in methods courses were correlated to student teaching grade, positive correlations were found with less significant ones for ratings. This may demonstrate the halo effect of ratings since many of the student teacher supervisors taught the student teachers in methods courses. Anderson (1965) conducted the only study which attempted to compare student teachers to certificated and experienced music teachers, and his findings are of dubious value. Therefore, one cannot ascertain whether music student teaching success per se is related to music teaching success.

3. Because of the interesting findings that personality and motivational variables contributed most to predictions of music student teaching success, it was decided to concentrate on these variables in more detail in this study. Morgan's study was most helpful in contributing to the design of this present study and will be discussed in more detail in Section Three of this chapter. However, in order to predict success one must first define a group of successful music educators and see if this group relates to a music student teaching group. The interesting nature of this problem provided the rationale for this investigation that it should concentrate on the music educators first and attempt secondly to cross validate any findings with a group of student teachers in music as well as another music educator group. In other words, the music student

teachers per se were not under investigation except for cross validation purposes.

4. It was recognized in the literature that a problem could be anticipated in the criterion domain. The most common criterion was some kind of supervisory rating. Because a similar criterion was needed for both the music educator and music student teacher groups, the investigator decided to use a composite rating as the criterion. However, the interrelationships of ratings varied as shown by Erickson (1954) and Morgan (1946). Anderson (1965) found high interjudge reliabilities among cooperating teachers and supervisors; yet, it could be hypothesized that these ratings were made from the same frame of reference. Therefore, five ratings were selected for investigation: pupil, cooperating teacher, supervising teacher, peer music student teacher, and self ratings. These would be appropriately weighted after analysis in order to derive a composite rating criterion, e.g. like Morgan's residual rating criterion.

5. Because many of the studies in the literature used student teachers from one specific institution, usually the degree granting institution, this investigator decided to try to get a cross section of music student teachers from six Illinois teacher training institutions and from the writer's university. This provided a multifarious music student teacher sample in terms of training, experience, musicianship, and background similar to those persons in the music educator sample. By using only University of Illinois music student teachers, a highly select sample (and probably a biased one at that) would be evolved. Therefore, the literature showed that music student teaching success could be predicted to some degree of accuracy, and it was decided to attempt to further clarify this area in this present study.

6. The studies in predicting teaching and music teaching success were generally positive depending upon the criterion selected which is the crux of the problem. In the general education area, teaching success can be predicted to some degree of accuracy. The studies in music have indicated that college

factors are not related to music teaching success, that musicianship is not related to music teaching success (probably due to the difficulty of a music test discriminating among musicians), that student teaching success may be related to music teaching success, and that personality, experiential, and background variables may be related to music teaching success.

Without discussing the specifics of the designs because these will be discussed in Section Three, this investigator decided to concentrate on the relationships and interrelationships of personality and motivational variables to music teaching success to the exclusion of other variables. Opinions from teachers who have left the profession tend to support this investigation in that the reasons they leave are not only for higher salaries but for other, deeper rooted motivational reasons. Studies by Lutz, Wink, Barth, and Fosse highlight the importance of personality variables to music teaching success.

The difficulty in designing a prediction study is in the selection of the criterion. Section Two emphasizes this important area.

Published tests in the area of music teacher education have not been consistently valid to be useful in predicting success. Therefore, these tests were not selected because the investigator felt that personality and motivation tests probably better define the dimensions relating to trait differences among teachers than do vocational guidance or interest batteries.

A further reason supported in the literature for concentrating on the personality and motivation variables useful in predicting music teaching success is offered by Cady:

> With so many avenues opened up and yet to be fully explored, it would be regrettable to let the work of Nohavec (Morgan), Snapp, and Watkins, for example, disappear without further nurture . . . replication among researchers is an honorable endeavor (Cady 1967:150).

Another reason is offered by Prince:

> In summary, replication of important research projects in music education is greatly needed to provide more secure evidence concerning vital questions . . . In contradistinction to the practice in historical, philosophical, and analytical research in music education of seeking unique topics for study, the practice of replication in experimental and some descriptive research in music education should be avidly pursued (Prince 1970:6).

It was with the above in mind that this research investigation was undertaken. In conclusion, it appears that technique, tools, and instruments are available, which, to some degree of accuracy, can predict college, student teaching, and music teaching success. This review of the literature encouraged the writer to pursue this area in depth.

SECTION TWO: TEACHER ACCOUNTABILITY CRITERIA

The literature pertaining to teacher effectiveness and success is immense numbering 2,000 or more studies since the 1920's. Fosse states:

> The crux of the problem caused by this bounty is that the researcher must: (1) develop (or adapt) a classificatory system for dealing with various criteria of "effectiveness" and classes of "predictive" variables and (2) present his review of the previous research in such a manner as to be representative and unbiased (Fosse 1966:73).

Naturally, the bias of the researcher is going to be present simply by the design he accepts for implementation of his research investigation. It is important at this point to mention a few of the present approaches used to evaluate music instruction in the schools and the problems presented by them. No research study can adapt a complete model to cover all the aspects of the music teaching act; however, an investigation can attempt to be as complete as possible within the normal limits of time and money.

One of the first difficulties in evaluating music teachers relates to the previous philosophical discussion. Without one accepted philosophy, no single test, battery of tests, ratings, process judgments, and so forth can be used for all music teachers. Lehman says, "Many persons are of the opinion that

BEST COPY AVAILABLE

the failure of music educators to define their objectives in precise and specific terms has been the greatest single obstacle to the meaningful evaluation of musical instruction (Lehman 1969b:427).

Today, basic decision making roles are being delegated to management consultants, systems analysts, cost accountability specialists, program planning and budgeting specialists, and administrators and supervisors within the teaching profession. Everyone is concerned with education be he taxpayer, legislator, parent, student, or combination of the above. Through legislative mandates enacted by state legislatures, new evaluation programs have been implemented. These may be Program Planning and Budgeting Systems (PPBS), Program Evaluation and Review Techniques (PERT), or a host of other approaches found in the literature. That persons are concerned with educational accountability is a fact. Today, a total of 32 of the 50 states have some type of law about educational accountability on their books (Killackey 1973:23). Many needs assessment programs are including local patrons as well as school staff members in all curriculum areas, the writing, implementing, and evaluating of the instructional program, state department and other "experienced" evaluators to make on the spot assessments of a school district's effectiveness, and plans for continued inservice training for educators. Music education must be included in assessment and accountability plans if funds are to be forthcoming.

Other groups interested in evaluation include the National Assessment Program supported by the Carnegie Corporation, the Ford Foundation, and the U. S. Office of Education, and administered through the Exploratory Committee to Assess the Progress of Education. Data are being collected for boys and girls separately in 4 geographical regions, northeast, southeast, middle west, and far west: and for 4 age groups: 9, 13, 17, and adult; for large cities, small cities, suburban areas, and rural areas; for upper and lower socioeconomic levels. While it will be unlikely that an individual would be asked to participate more than once in his lifetime, items included for evaluation are (1) those considered important by experts in the respective fields, (2) those accepted by educators as tasks of the schools, and (3) those deemed desirable by leading lay citizens. Music is one of the fields included in the National Assessment Program. Furthermore, under the Advanced Placement Program of the College Entrance Examination Board (CEEB), many students will be able to complete college level courses in music while still in secondary school (Lehman 1969a: 27-28).

This discussion has concentrated on the objective aspects of music education. Certainly music educators will be required to objectively evaluate their students and themselves in order to continue to meet school district and state requirements. As the following discussion will show, there are many difficulties with using pupil gain scores, pupil norm scores or centiles, or some other kind of pupil criterion as an indicator of music teacher effectiveness or program effectiveness. Yet, this is one important area that must be considered in any evaluation model offered for public consumption. These types of measured behaviors may be classified as preordinate measures in that they call for prespecification of program and instructional outcomes. Achievement, performance, and observation tests and checklists are selected and developed to provide evidence that prespecified goals were or were not achieved. This preordinate approach depends upon a capability to state the important purposes of education in terms of pupil behavior and a capability to discern the accomplishment of those purposes.

Stake (1973) criticizes the preordinate approach because it is not sensitive to ongoing changes in program purpose, nor to unique ways in which students benefit from performing in art media or from encountering artistic expression, nor to dissimilar viewpoints that people have as to what is good and bad. He recommended the "responsive" evaluation approach as a time tested, thorough, and sensitive evaluation procedure for the purposes of most educators and citizens about the arts in our society (Stake 1973:1).

Stake recognizes the fact that there is no escaping evaluation in the world ahead but he emphasizes the importance of evaluating with sensitivity so that the results are useful. In the relation of evaluation to purposes and audiences, Stake raises the following points about why people expect evaluation to have many different purposes:

1. To document events
   When did the parents become interested?
   When did the aims of the program change?
2. To record student change
   Are these students of contemporary art aware of similarities in expression across different media?
   Do the students like poetry more than they did?
3. To aid decision making
   Should community artists be asked to contribute time?
   Should the CEMREL materials be purchased?
4. To seek out understanding
   Why does this band program result in excellent student participation?
   With what kinds of students does that teaching style work?
5. To facilitate remediation
   How can we honor the aesthetic values of the students, persuade them to honor ours?
   How can we make the self-study program a better program (Stake 1973:4-5)?

Each of the above questions deals directly or indirectly with values, the values of an educational program in the arts. Therefore, each is a legitimate evaluation question. He says:

> I think of evaluation as a service activity. I expect it to be useful to specific persons. The evaluation will probably be more useful if the evaluator understands the interests and the language of his audience. A substantial portion of the time of an evaluation study may be allotted to informal inquiry into the information needs of the persons to be served by the arts program. The program is supposed to do things for people, not necessarily the same people. The evaluator should have a good sense of whom he is working for (Stake 1973:6).

Stake believes strongly that most educational evaluators are merely concerned with the instrumental values of education: the idea that good education results in specific outcomes such as performance, mastery, ability, and attitude. This is wrong according to Stake for the long run because educational results may be delayed for years and therefore "It is sometimes the purpose of the program staff or parent to provide artistic experiences, and training, for the intrinsic value alone" (Stake 1973:6). He continues by comparing preordinate and responsive evaluation:

> Preordinate plans for the evaluation of educational programs emphasize (1) statement of goals, (2) use of objective tests, (3) standards held by program personnel, and (4) research-type reports. Responsive evaluation is an alternative, an old alternative. Responsive evaluation is what people do naturally in evaluating things. They observe and react. What is new about it is a technology, designed to overcome some of the defects of this natural behavior for formal evaluation purposes (Stake 1973:7).

Therefore, an educational evaluation program is responsive if it (1) orients more directly to program activities than to program intents, (2) responds to audience requirements for information, and (3) if the different value perspectives present are referred to in reporting the success of the program. Stake prefers to discuss and write about issues in his evaluation reports. Furthermore, he emphasizes the need to write the evaluation report for the specific public or reader. A highly technical report might suffice for the school's research director but not for the school board. The crux of Stake's solution to the evaluation dilemma in arts lies in the portrayal, the emphasis on verbal description of the hows and whats of a particular program.

Edwards (1973) emphasizes the importance of the fundamental relationships of two basic tasks to that of producing an audience appropriate description of the curriculum:

> 1. He (the evaluator) must determine the role that comparison is to play in the evaluation effort. All evaluation ends in some form of comparison, albeit a subjective standard. The question is not whether comparison will be made, but rather at what point comparison will enter the decision making process; and, when it does so, with what comparison will be made. The way in which this task is treated sets an important tone for the evaluation by establishing what form the data must take in order that they can be meaningfully interpreted by the consumer in the context of whatever decisions are to be made.

> 2. He must determine precisely what information will be utilized in this framework by specifying what dimensions of the program will be studied and by what means (Edwards 1973:2).

Edwards feels that simply getting an opinion from the arts consumer is not enough. Behavioral objectives do provide a simple answer to the question of what to look for in a program of instruction. Likewise, norm referenced tests facilitate the task by providing psychometrically defensible variables. "With behavioral objectives, norm-references instrumentation, and a well-designed experimental model, the evaluation effort would seem to be shored up against all attacks; there are objectives to justify the choice of variables, statistical descriptors that will defend the quality of their measurement, and a sound basis for defending the validity of the experiment and the resulting conclusions" (Edward 1973:3-4).

He feels that the first step is to make the consumer aware that a curriculum can be evaluated with regard to the degree to which it fulfills some set of predetermined objectives but that this is only one way to do so. The consumer must also be aware that he does have some choice in determining the standard for individual and joint success with regard to a given program and that this standard can be either relative or absolute. Experimentation provides both advantages and limitations in decision making. "And most importantly, both evaluator and consumer must understand that unless these decisions can be arrived at with the agreement of both parties, the evaluation effort will be less than satisfactory" (Edwards 1973:4).

Therefore, a consumer oriented philosophy is urged regarding arts evaluation. In measuring behavioral objectives, Edwards stresses the importance of utilizing criterion referenced rather than norm referenced tests. Criterion referenced tests refer to measuring the goals or objectives arts courses or programs hold in common rather than comparing individual scores to national norms which may or may not be appropriate for the local situation. Criterion referenced tests aid the teacher in arriving at decisions about how well he reached his own and the class's objectives and goals. Edwards suggests that when comparisons can be made on a large enough scale, classrooms or teachers should be the sampling unit rather than individual students.

> The decision to use norm or criterion-referenced tests already alluded to is another about which the evaluator must be in a position to provide advice to his arts oriented consumer. This problem also bears close ties to the overall questions of audience appropriateness and design. Many audiences (particularly students, teachers, and writers) will often instinctively lean toward criterion-referencing of student outcomes if they are made aware that it exists. To them how students are doing is most easily explained and understood when comparison is made to how they should be doing, not the "doings" of other individuals or groups (Edwards 1973:21).

That there are many difficulties in arts evaluation is true. However, certain aspects of the music education curriculum can be objectively measured and reliably and validly evaluated. This being so, these evaluations can be used to determine teacher effectiveness. Evaluation in previous discussions has often referred to evaluating the success of a curriculum or a new methodology of teaching. In the context of this present investigation evaluation refers to evaluating the teacher and finding a model best suited to this purpose. Because of time limitations, descriptions or reports would have been too expensive and difficult to gather considering the location of participants. Therefore, it was decided to use current thinking in arts evaluation and to measure what could be measured feasibly. These were in two main areas: pupils' music achievement of very limited music behavioral objectives using a norm referenced music achievement test in a criterion referenced technique, and ratings of music teaching success which could be considered to be subjective evaluations.

Colwell (1970) emphasizes that evaluation is an ongoing process, circular in nature, with each stage contributing to the next one sequentially. The process centers around musical problems because these are the content of the music curriculum, and the last step of the process contributes to the first step by helping to plan the future musical experiences; in other words, to assist decision making. This circular process of planning and teaching music, collecting evidence and data, processing and interpreting the data, and disseminating and using the information applies specifically to the teacher as the evaluator.

> The process is slightly different if the evaluator is someone from outside the classroom - supervisor, consultant, or administrator. In this instance, the first step of planning and presenting learning experiences is replaced by planning the evaluative tools and procedures; interpretation of data may require the help of the teacher and others. The responsibility of using the information will again fall to the teacher; therefore his information must be complete and his understandings of it adequate (Colwell 1970:22-23).

Colwell warns about the danger of using evaluative devices to select objectives for the musical learning experiences:

> The objectives themselves should not be selected on the basis of information gathered through evaluation. Objectives are based upon, and draw their strength from, a philosophy of music education, a theory of what elements are important for well-balanced music education at any given level of growth (Colwell 1970:23).

In other areas of music, musicians have been alert to evaluation and measurement; the performing musician and the musicologist. However, music educators have been noticeably lack in developing evaluative skills:

> Why are music educators careless and seemingly unconcerned about the achievement of goals? Perhaps because they so often deal with children rather than with people old enough to have mastered musical skills and knowledges that can be challenged. In other public school areas, however, teachers are concerned with real progress rather than with pleasant experiences (Colwell 1970:26).

With the above views in mind, it was felt that participating music educators could select a few limited goals which could be objectively evaluated. By limiting their selection to goals universally considered to be important in music education, it was felt that the philosophical requirements would be met.

A recent review of studies of stability of teacher effectiveness (Rosenshine 1970) could locate only four longitudinal investigations which included stability coefficients reflecting teacher consistency across time in producing student gains. The stability coefficients were generally low in these four studies. Recently, Brophy (1973) in a three year longitudinal study raised the following questions:

1. How stable is the effectiveness of typical teachers?
2. Are some teachers more stable than others? (Brophy 1973:245).

His study is crucial to any research investigation deriving a teacher effectiveness construct and attempting to predict potential teaching success. If teachers vary, especially superior teachers, from year to year in their abilities to produce learning in children, the question of prediction would be moot.

He employed in his study teachers teaching in the second and third grades in a school district: 15 Title I schools and 35 non-Title I schools. The Metropolitan Achievement Tests provided pupil data for computation of residual gain scores. He used only the teachers who taught the same grade for 3 consecutive years and had classes of 14 or more students since he felt smaller classes were likely to be a non-random sample. Four sets of data were collected: gain scores for second (N=54) and third grade (N=51) teachers in non-Title I schools and gain scores for second (N=34) and third grade (N=26) teachers in Title I schools. Correlational analyses of these four sets (mean residual gain class scores) were used to investigate consistency across the subtests within year and stability across the two sexes and the three years within subtest. He separated the pupils into sex groups because girls have a tendency to gain more during these years of elementary school (Brophy 1973:247).

He concluded:

> In summary, the findings indicate that, at least in grades two and three, teachers who are consistent in their overall relative effectiveness can be identified. These teachers are about equally successful with boys as with girls, and about equally successful across three years in producing gains on the four or five Metropolitan subtests on which data are available. Other teachers show a more complex kind of consistency, such as producing high gains in language arts and low gains in math, or vice versa. Observational studies of these consistent teachers, done in the naturalistic setting as they carry out normal activities, should yield greater payoff than the kinds of teacher effectiveness

> research done in the past, since process measures of their behavior should be both more reliable and more valid than measures of an unselected group of teachers. Ultimately, such research should yield process-product linkage data that would provide a rational and valid basis for constructing accountability criteria (Brophy 1973:251).

In regards to the gain scores themselves and probable sources of error in the residual gain technique, Brophy concluded:

> Given the probable masking of some consistency by the militations of the data and given the many uncontrolled factors operating to affect a teacher's effectiveness during a given year, it seems likely that the present data, to the extent that they may be in error, are underestimates (Brophy 1973:251).

He feels that indiscriminate use of pupil gain or general achievement tests for assessing teacher accountability should be avoided until the sources of instability become known and eliminated or controlled. This procedure, otherwise, could be unfair to the participating teachers.

Brophy showed that teaching effectiveness is moderately consistent. Some teachers were low in their first year and then showed a strong consistency their last two years indicating improvement in teaching skills. The strength of this study, one of the few in the literature, supports the general thesis that when attempting to define teaching effectiveness and success, as broad a sampling of the criterion domain should be made, e.g. his suggestion for process-product measurements. Furthermore, he was able to show that the residual gain technique was useful in formulating teaching effectiveness constructs.

This present investigation decided to investigate the criterion domain of music teaching success within normal limitations. An objective pupil gain measurement required a standardized music achievement test, but, in order to make it valid for music classroom use (criterion referenced), music teacher agreement on certain selected music behavioral objectives or goals was required. Chapter Three explains this procedure in detail but agreement was obtained so that the advice of Stake, Edwards, and Colwell could be followed. Furthermore, since process variables appeared to be important to the teaching enterprise, the decision to use a variety of rating evaluations was made.

The following discussion emphasizes the importance of the gain versus rating versus process controversy in the literature. There appears to be a cyclical nature to evaluating teacher effectiveness. The early studies investigated both pupil gain scores and ratings. The general consensus was that gain scores or product evaluations were important and a considerable amount of research was undertaken. Since the 1950's and the research of Ryans, Flanders, and B. O. Smith in particular, interest was directed toward process evaluations. The idea here is that if the classroom climate is appropriate for instruction, learning takes place. Numerous rating forms and checklists were developed to evaluate these process variables. Many studies emphasized the environmental conditions such as the democratic versus the autocratic teaching behavior, teacher centered versus learned centered behaviors, and a host of these kinds of variables.

Recently, interest has returned to product evaluations primarily due to the prominence of the behavioral objective movement. These behaviors, stated in terms of pupils' expected behavioral learning outcomes, can be objectively measured by tests. Pupil assessment became important and types of techniques for treating gain scores were evaluated. The data treatment techniques in the early part of the century were primarily partial correlational analyses. Accomplishment Quotients were developed using intelligence and previous learnings as covariates. The residual variation left after partialling resulted in the Accomplishment Quotient figure which was an attempt to be as precise as was the Intelligence Quotient. Today, the techniques are similar but more refined. Two generally acceptable methods have been used extensively in the literature: analysis of covariance (ANCOVA) and residual gain scores.

McNeil and Popham criticize the process product studies because of investigators' failures to use appropriate dependent measures. They feel that process product studies could be improved by: (1) paying more attention to

BEST COPY AVAILABLE

prior learnings of students and by increasing the number of predictor variables (covariates) to include socioeconomic status, type of school, and so forth, (2) accounting for stability of teacher behavior across occasions, (3) collecting evidence of the teacher's effectiveness from one year to the next (more studies like Brophy's investigations), (4) designing studies in which teacher behavior is studied over a short period of time such as with the microteaching technique, and (5) studying disordinal interaction, teaching behaviors which produce results with one type of learner but not with learners of different characteristics or traits (McNeil and Popham 1973:226). While these suggestions are most worthwhile, research in teacher education appears to have a long way to go before satisfying the above requirements.

In the present study, a considerable amount of effort was expended in order to determine the best possible approach to use because as will be seen, the criterion variables selected for the dependent measures significantly affect the research findings of any study. Discussed first is the gain versus rating dilemma, second pupil measurement, and third ratings of process as well as teaching effectiveness.

Gain versus Rating Dilemma

In the period 1924-40, A. S. Barr and others conducted a series of research investigations from which three conclusions were drawn:

> 1. The fact that 2 or more observers observing the same teacher simultaneously may disagree in the quality of teaching observed was reaffirmed;
>
> 2. Good teachers cannot be separated from poor teachers in terms of specific teacher behaviors (There is an appropriateness aspect to teacher behaviors that must be taken into consideration; and
>
> 3. The evaluation of teaching can be objectified through the use of teacher and pupil behaviors and operational definitions of the personal and professional prerequisites to teacher effectiveness (Barr 1961:ii).

In a report about 75 or more doctoral studies summarized by A. S. Barr in 1961, two criteria of teaching success were used: (1) efficiency ratings of one sort or another, and (2) measured pupil gains. "Over-all, general ratings of teacher effectiveness have been shown to be, under current conditions, exceedingly unreliable" (Barr 1961:8). He also states that the use of measured pupil gain as a criterion of teacher effectiveness also presents very real difficulties because each teacher chooses his own purposes, means, and methods of instruction. The tests used in developing the pupil gain criterion will have varying degrees of operational validity except as the teachers agree to pursue certain stated objectives which can be defined sufficiently well to provide like meanings to all participants in any research study. "A second difficulty arises out of the fact that, notwithstanding over a half century of effort, many of the outcomes of learning and of teaching are poorly or inadequately measured" (Barr 1961:8). He further states that tests measure effects but no causes. The sources of the effects observed are not readily ascertained even under carefully controlled experimental conditions.

Barr, Torgerson, and others (1935) employed some unique approaches to criterion development. Five criteria were employed: (1) gain in pupil achievement as measured by the Stanford Achievement Test, and gains being computed as (a) gain in the total raw score, (b) gain in the arithmetic score, and (c) gain in the Accomplishment Quotient; (2) a composite of scores on seven rating scales; (3) a composite of nine measures of qualities commonly associated with teaching success; (4) a composite of six tests of teaching ability chosen from composite three; and (5) a composite of all 19 variables combined in various ways. The major conclusions were: (1) the coefficients of correlation obtained between 10 selected measures of teaching ability and gain pupil achievement are uniformly low; (2) the coefficients of correlation between the 19 variables employed in the investigation and the 5 composite criteria provide conflicting evidence;

(3) a composite of the total pupil gain score on the Stanford Achievement Test, the Torgerson Diagnostic Teacher Rating Scale, and the Knight Aptitude Test yielded a multiple correlation of .70 with a composite of all measures employed as criteria; (4) 14 of the measures had a forecasting efficiency superior to general merit ratings; and (5) when the Torgerson Diagnostic Teacher Rating Scale and the Knight Aptitude Test were employed to predict ratings in 5 categories representing A, B, C, D, and F letter ratings, 64 percent of the predictions fell within the correct criterion group, 32 percent were misplaced by one group, and only 4 percent were misplaced by two groups (Barr 1961:16).

> Based upon data such as these and considering our present state of our knowledge, many persons believe that if one desires an overall criterion of teacher effectiveness, probably the safest procedure is to employ a variety of measures, all possessing validity from some particular point of view, applied and evaluated by more than one person, and based upon studies of the teacher over a considerable period of time. If one does not desire an overall competency evaluation, the several components may still be considered separately. If one examines the correlations from this point of view it can be observed that none of the correlations with the pupil gain criterion are high with about as many negative correlations as positive ranging from -.32 for the Giles Teacher Scale to .23 for the Wood Health Scale . . . When a composite of nine measures of qualities thought to measure teaching ability was used as the criterion all of the correlations were around zero: .08 for pupil gain, and -.06 to .12 for the six teacher rating scales . . . Very clearly the different criteria gave different results. One can make differential predictions but the effectiveness of a prediction depends upon the criterion (Barr 1961:16).

The importance of Barr's findings is in the relationships of experimental variables to certain criterion variables. Depending upon the criterion, one may be able to produce different experimental variables correlating to it. Thus, criterion validity assumes the major importance in any prediction study. If one defines success in terms of pupil gains, different predictions from the same set of experimental variables may be formulated than if one were to use some type of rating or process criterion.

Lins (1946) employed three kinds of criteria in his study; supervisory ratings, pupil ratings, and residual pupil gain scores. The supervisor rating criterion was a composite of 5 ratings; two by members of the Department of Education, University of Wisconsin, one by a representative of the State Department of Public Instruction, one by the principal or superintendent under whom the teacher taught, and one by a member of the department of educational methods under whom the teacher had practice taught. The ratings were all made on the Wisconsin adaptation of the M-Blank of the Evaluative Criteria plus certain other instruments. The ratings were preceded by almost one full year of weekly discussions among the members of the team responsible for the collection of the data and each teacher was visited by a trained observer at least once for more than a single class period, usually a half school day, in which the teacher was observed at work and interviewed. The pupil ratings were accomplished according to a carefully designed plan and under the direction of a single person who visited each school and secured the evaluation by anonymous ballots. The residual pupil gains were the discrepancies between the actual gains made upon certain standardized subject matter tests and a predicted pupil gain based upon a four variable prediction equation derived from average gain scores, the pretest score, mental age, and an intelligence quotient (Lins 1946:2-60).

One of the most important findings was the low intercorrelations among the three criteria: pupil rating and composite M-Blank (.279), residual pupil gain and composite M-Blank (.193), and residual pupil gain and pupil rating (.055). Lins was concerned with predicting teacher effectiveness when viewed by a variety of persons, under varying conditions, and where more than one sort of criterion was employed. Even with low correlations, Lins was able to obtain a multiple R of .74 with the residual pupil gain, .72 with a composite of supervisory ratings, and .60 with pupil ratings. These findings are in substantial agreement with those of Barr and others.

These findings suggest that the criterion domain reflects what the researcher desires and the more measurements taken and techniques used the better the predictions may be to some extent. Anderson (1954) studied the criterion itself and based his analysis upon data gathered through the use of the following data gathering devices:

1. The principals' ratings of the teacher upon the Wisconsin Adaptation of the M-Blank (M).
2. A rating of the teacher by pupils on an adaptation of the Bryan Reaction to Instruction Questionnaire.
3. A teacher self rating upon a special form (SE).
4. A rating of the teacher by his peers upon a special form (Pe).
5. A state supervisory rating (Ax).
6. Pupil score on an initial administration of standardized achievement tests (P.T.).
7. Final scores on selected standardized test (F).
8. Residual pupil gain (AG).

Anderson concluded:

1. No correlations appreciably different from zero were found between the evaluations of teachers made through the several ratings and pupil achievement as measured in this investigation.
2. The achievement tests employed in such investigations must have curricular validity and the pupil gains must be large enough to be statistically significant.
3. Most ratings are very subjective and the intercorrelations among them low (Anderson 1965:41-77).

Anderson supports the conclusions of Lins and Barr and others. The criterion domain appears to have separate factors which are basically unrelated and uncorrelated. Erickson (1954) in a factorial study of teaching ability found no general factor but rather three measures, each of which might be used separately in a program of differential prediction. All of these above findings underline the difficulties involved in criterion development.

Barr concluded his survey of the literature by stating:

> 1. The two general criteria of teaching efficiency, pupil change and rating of teaching, in almost all studies tend to become markers as definers of factors . . . There is nothing in the factorial method that should artificially produce this result if the criteria weren't logical definers of the domain.
>
> 2. Pupil gain factors bear no relationship to rating factors.
>
> 3. There are at least two kinds of pupil change occurring under the influence of the teacher . . . that these might be representative of informational and non-informational changes.
>
> 4. In general, ratings of teaching efficiency were undimensional. However, Erickson's data indicated that self-ratings might be an additional criterion factor.
>
> 5. Of the many correlative variables in these studies, only a few, and these not consistently so, bear relationship to the pupil change factors. Among these were: the teacher's intelligence, attitudes, ability to recognize, diagnose, and correct pupils' emotional problems, and the teacher's knowledge of the subject matter. It should be emphasized that these variables were neither consistently nor highly related to pupil change factors. In general, objectively measured personality characteristics were not found on pupil change factors.
>
> 6. The rating factor remains enigmatic. In general, few correlative variables were related. However, the evaluation of teacher's future proficiency by interview seems factorially better than by evaluation of autobiographical information (Barr 1961:68-69).

Hellfritzsch (1945) in a factor analysis of a number of measures frequently used in research investigations found that supervisory ratings were not related to pupil gain, knowledge of the subject, mental ability, or social and emotional adjustment, but were divorced from any of the objective measures employed. Therefore, the literature is fairly consistent in that objective measures such as pupil gain scores are not significantly related to subjective measures such as supervisory or composite ratings.

When selecting criterion variables, the investigator may predetermine his findings without really intending to do so. If only one domain is represented, such as a rating criterion, the variables that predict that criterion only hold for the type of rating instrument used in the study. If one wishes to generalize his findings, he will find the task difficult. From a perusal of the literature, the dependent variables appear to be most critical, and the selection of these most important for any type of teacher effectiveness and prediction study. Before one can predict success, the definition must be useful, replicative, and valid.

Pupil Gain Scores

> One of the most elusive targets in the history of educational research is a valid index of teacher effectiveness. Since the turn of the century literally hundreds of investigations have probed the question of teacher competence assessment and most of them have produced little, if any, significant progress (Popham 1971:105).

Popham emphasizes the need to focus research activities toward identifying teachers who could produce superior learner growth, leaving aside the question of how such improvements were caused or brought about. He emphasized that while researchers have employed the criterion of learner growth as an index of a teacher's proficiency, these efforts have used norm referenced standardized achievement tests rather than criterion referenced approaches to test construction. Furthermore, there has been a marked increase in research employing classroom observational techniques and have too frequently attended to classroom process variables without consideration of resulting modifications in learner behavior. The major efforts have used the Flanders' Interaction Analysis procedure, Smith and others' Categories for Strategies of Teaching, and Ryan's Classroom Observation Record (Popham 1971:106).

Popham designed a procedure for assessing teaching competence involving the use of performance tests of teaching proficiency. The procedure for using such performance tests requires that an instructor be given the objectives and resource materials well in advance of instruction, then he devises a sequence of instruction suitable for accomplishing the objectives using whatever instructional procedures he wishes. The teacher would be obliged to accomplish the prespecified objectives but would have freedom to choose instructional procedures which seemed likely to achieve these goals (Popham 1971:107). Three specified courses of study - Social Science Research Methods, Basic Power Supplies, and Carburation - were taught by experienced certificated secondary school teachers and nonteachers in San Diego and Orange County, California, schools. An honorarium of $25.00 for social science and $50.00 for electronics and auto mechanics teachers and nonteachers was paid to each participating subject. Popham piloted his pre and posttests of which only social studies classes were not pretested due to time limitations.

Using ANCOVA procedures with the pretest and students' expressed interest in the topic as covariates and the class as the instructional unit rather than individuals, Popham found no significant differences between the abilities of teachers and nonteachers to promote learner attainment of prespecified instructional objectives. "Results of all three validation replications failed to confirm the prediction that experienced teachers would promote significantly better achievements of given instructional objectives than would nonteachers" (Popham 1971:114).

The measuring instruments appeared to be valid and yielded satisfactory criterion referenced validity standards. Furthermore, the teachers were at an advantage because they were familiar with the school setting, knew the pupils, and were skilled in classroom management and disciplinary techniques. While the nonteachers taught, because of legal requirements, a substitute teacher was unobstrusively located in the rear of the room. Popham concluded, "Although there are competing explanations, such as insufficient teaching time, the explanation seems inescable probably is the following: Experienced teachers are not particularly skilled at bringing about prespecified behavior changes in learners" (Popham 1971:115). He continued:

> When it comes to a task such as that presented by the performance test in which they must promote learner attainment of specific instructional objectives, perhaps most experienced teachers are no better qualified than a person who has never taught. To realize why this might be so, one needs only to speculate on the typical intentions of most public school teachers. They wish to cover the content of the course, to maintain classroom order, to expose the student to knowledge and so on. Rarely does one find a teacher who, prior to teaching, establishes clearly stated instructional objectives in terms of learner behaviors and then sets out to achieve those objectives. Only recently, in fact, do we find many teachers who are even familiar with the manner in which instructional objectives are stated in measurable form . . . The general public, most school systems, and professional teachers' groups rarely attach special importance to the teachers' groups rarely attach special importance to the teacher's attainment of clearly stated instructional objectives (Popham 1971:115).

While Popham raises the question that a research design attempting to validate a performance test of teaching proficiency by contrasting the accomplishments of teachers and nonteachers may have been ill conceived, the fact that this validation scheme was injudiciously selected does not mean that the performance test approach is unworkable, nor that such tests cannot be validated.

> It now appears, in light of the grossness of the measurement devices likely to be available in the near future, that we shall be pleased even if the performance tests are suitable for use only with groups. In other words, it will be a sufficient advance to develop a reliable group criterion measure which could be used in myriad educational situations such as to assess the efficiency of teacher education programs (Popham 1971:116).

For the purposes of this present study, Popham's views were accepted for one aspect of obtaining music teacher accountability criteria. A type of criterion referenced music achievement testing procedure was used in order to gather data for the pupil residual gain scores. The following discussion highlights the techniques used in the literature to treat gain or change scores and are, therefore, pertinent to this investigation.

Barr (1938) emphasized the use of two assumptions which must be made about using gain scores as teacher accountability criteria before the word "accountability" became fashionable:

> If this approach (gain scores) is made to the measurement of teaching ability, it should be remembered that certain important assumptions have been made: (1) We possess at the present time adequate measures of the major changes produced in pupils . . . (2) Those factors other than teaching ability can be controlled, equated, or otherwise held constant as in experimental research. Teaching is only one of several factors conditioning the changes produced in pupils. If the gains in test scores for one teacher are to be compared with those of another, those factors other than teaching ability affecting the products of learning must be controlled, equated, or otherwise held constant, as in experimental research. This latter condition is sometimes not possible in field research (Barr 1938:472-473).

Barr further states that "the use of test scores for evaluating the efficiency of teachers is an exceedingly delicate process . . . Although the method is theoretically sound, all told, more harm may be done than good, except as the method is applied with great care" (Barr 1938:473). Many of the studies at Barr's time used part and partial correlational techniques in order to derive Accomplishment Quotients. Multiple regression techniques were used, but before the advent of computers, these techniques were considered to be impractical for most situations because of the time consumed in ob[illegible] variable weights by hand calculations.

Barr summarized the difficulties of using gain scores by [illegible]ng, "errors are introduced into the Accomplishment Quotient: (1) because of the use of [illegible] fallible measures of capacity and achievement - the probable error[illegible] tients composed of fallible measures are larger than the probable err[illegible] the measures themselves; (2) because general intelligence and the ability to perform in the different school subjects is not perfectly correlated; (3) because gain scores made at different points of the accomplishment curve are of unequal value; (4) because of extraneous factors such as health, home environment, and competing interests; and (5) because teachers and courses of study parallel the content and objectives of tests used with varying degrees of faithfulness" (Barr 1938:476).

BEST COPY AVAILABLE

The major assumptions at that time for using the AQ technique involved the following:

> 1. That the capacities measured by tests of intelligence correlate and correlate equally well with the ability to achieve in the different school subjects.
>
> 2. That intelligence, except for teaching ability, is the chief factor conditioning the achievement of pupils, or that this factor is highly or perfectly correlated with other factors, or that these factors are insignificant.
>
> 3. That the units of gain are linear, or equal for different levels of intelligence and attainment (Barr 1938:475-476).

The same things that were causing concern for Barr and his associates are still affecting gain scores today. Any attempt to measure pupil gain is subject to the fallibility of the test instrument. If one student were measured many times on a test, the mean of these measurements ordinarily approaches a limit as the number of measurements increases. This means that the errors of measurement are being averaged out. When these errors of measurement reach this limit, the limit is called the "true value." In any situation, the person measured behaves at least to an adequate approximation as if a true value actually existed even though this true value cannot be operationally determined. This makes true value a useful practical value (Lord 1963:25-26).

The difference between true and observed values or scores on a test is treated as an error measurement. A seventh grader who scored 44 on Music Achievement Test I would have a true score between 39.79 and 48.21 approximately 68 percent of the time and between 35.58 and 52.42 approximately 99 percent of the time (Colwell 1969:66).

Posttest minus pretest measures of learning produce difference, change, or gain scores which almost certainly will have a negative correlation with the pretest scores upon which they are based.

> This is considered a defect of such "difference" scores (gains) when we have reason to believe that amount of learning should not necessarily correlate negatively with pretest status . . . An alternative method is to measure learning or change (gain) by fitting a straight regression line to the pretest and posttest

> achievement test data and taking the deviation from the regression line (errors of estimate) measured along the posttest axis (Glass and Stanley 1970:182).

This deviation is called residual gain.

Lord discusses a widespread fault in speech and thought to substitute observed value for the true value. This common sense approach leads to many misconceptions and false interpretations of data (Lord 1963:26-27).

True change or gain is a residual standing for whatever true variance in the posttest cannot be accounted for by variance in the pretest score (Bereiter 1963:15). True change can be predicted and estimated from observed values by ordinary multiple regression analysis (Lord 1963:28). The predicted or estimated values have a smaller range than the observed values because the observed value scores contain errors of measurement that tend to inflate their range over what it would otherwise be. Estimated values obtained from regression analysis show less variability than the values being estimated. "This is an inconvenient property of such estimated values, arising from the regression effect, but this is the price we pay for having estimates that minimize the squared error of estimation" (Lord 1963:30).

Cronbach and Furby (1972) suggest that gain scores are rarely useful no matter how they may be adjusted or refined and suggest that investigators who ask questions regarding gain scores would ordinarily be better advised to frame their questions in other ways (Cronbach and Furby 1972:80). Lord has changed his mind in recent writings:

> In the writer's opinion . . . there is no logical or statistical procedure that can be counted on to make proper allowances for uncontrolled pre-existing differences between groups. The researcher wants to know how the groups would have compared if there had been no pre-existing uncontrolled differences. The usual research study of this type is attempting to answer a question that simply cannot be answered in any rigorous way on the basis of available data (Lord 1972:305).

Marks and Martin (1973) further substantiate Cronbach and Furby's and Lord's statements:

> As a general rule of thumb, the investigator computing true gain estimate should only employ test forms with reliabilities in excess of .85 and especially so if the true gain-initial true score correlation is expected or found to be .70 or less. To do otherwise would probably result in an unacceptable large confidence interval on (predicted gain), and an all but useless (predicted true gain score) (Marks and Martin 1973:190).

Because the Music Achievement Tests I-IV have reliabilities sufficient to substantiate the use of the residual gain technique, at least in the opinions of Marks and Martin, this study decided to use the technique to arrive at a change or gain score as a teacher accountability criterion. The following studies refer to specific differences between the residual gain technique and ANCOVA.

In previous writing, Lord (1963) suggested using a multiple regression approach in order to obtain a residual gain score, partialling out pretest, regression, and test error effects. While the above statement has value regarding experimental research, one might be able to use gain scores as theoretical constructs of teacher accountability.

Cronbach and Furby suggested an approach for computing an adjusted gain score which included the multivariate nature of change within an analysis of covariance design. They felt that learning or growth was multidimensional and that demographic, experiential, and organismic information (which they call W and Z variables) can and should be considered in this procedure for adjusting posttest scores both in ANCOVA and residual techniques (Cronbach and Furby 1972:73). They reduce the emphasis on the special role of X (the independent variable, usually the pretest score) as precurser of Y (usually the dependent variable, posttest score), and regard the whole WX set as a vector describing the person's true scores for pre and posttests as Wx, Xx, Yx, and Zx. They argue that even if treatment groups are not found randomly, ANCOVA can be used; and it is advisable to regress the covariate toward the mean of the treatment group before entering it in the analysis:

> If there is W information as well as X, it also contributes to the estimate. So does Y and Z information. Here is a paradox: a proposal to use the posttest score to estimate the pretest true score which will then be used to adjust the posttest scores! The crucial point is that the estimator of the covariate is determined from within-group data (Cronbach and Furby 1972:78).

They feel, however, that even this technique is merely no more than a palliative.

Dyer-Linn-Patton (1969) used four approaches which they called "discrepancy measures" to evaluate the effectiveness of school systems. They deplored the common tendency to compare the average achievement test score of students in a given system with some sort of national average or norm and to assume that the discrepancy between the two averages constitutes a measure of the educational effectiveness of the school system (Dyer-Linn and Patton 1969: 591).

Two of their approaches were of value to this research project. Method-I computed the pupil individual gain scores for each school system in order to derive an adjusted mean score while Method-M averaged pre and posttests for each system and used these means to compute residual gain scores for each school system. After reanalyzing the data of Dyer, O'Connor states that the Method-M approach is superior:

> . . . Method-M is preferable when computing residuals for groups. While Method-I residuals may be more reliable, they are biased when they are correlated with both the inputs and the predicted outputs. They should be uncorrelated with both. In fact, the Method-I residuals are more reliable because they are systematically biased (O'Connor 1972:88).

Method-I is similar to analysis of covariance (ANCOVA) while Method-M is the residual gain technique. O'Connor believes that unique to the residual approach is the selection of appropriate input measures:

> In the long run, the selection of the inputs may turn out to be more critical than the selection of the output measures. There are some indications that the use of socioeconomic variables as the inputs may produce a set of residuals uncorrelated with the set of residuals produced by using

> prior achievement as the inputs. Probably both kinds of variables should be used in the same set of equations (O'Connor 1972:89).

O'Connor emphasized the point that group means should be used to compute residuals for comparing groups and that it is preferable to have all groups similar in size; otherwise, the residuals may vary greatly in their reliability. O'Connor recommends using the residual technique with dummy variables (W and Z information). In private communication to the writer, O'Connor further highlighted his beliefs.

> If your subjects have been randomly assigned to the teacher, there should be only chance differences between the classes initially and the residual and ANCOVA techniques will produce very similar results; the correlation between the two sets of "teacher effects" may be in the high nineties. However, if the subjects were not randomly assigned, the two techniques may produce very different results, particularly if there are large differences in the groups on any of initial status measures. Even so, the correlation between the two sets of effects could be in the seventies or eighties (O'Connor 1972b:1).

O'Connor prefers the residual approach, Method-M of Dyer and others, and states why:

> As I reported in my article, the Method-M or residual approach is preferable to the Method-I or ANCOVA approach because the Method-I discrepancy measures are biased. The reasons why Method-I failed are complicated; regression toward the mean and measurement error are part of the problem in a fuzzy sort of way, but there are other difficulties as well. The residual approach seems to particularly alleviate some of these problems (O'Connor 1972b:2).

Research by Cronbach and Furby, Dyer and others, Marks and Martin, Lord, and O'Connor hint that gain scores may be useful as teacher accountability criteria but under carefully controlled circumstances. Therefore, a hybrid residual gain technique was adopted in this study as a good approach to use as found in the literature.

The early research by Barr and his associates also used a type of predicted gain score. Their technique, the Accomplishment Quotient, did not adequately control for variations in teacher objectives and goals because, in general, the same standardized tests were used for the whole sample. Today, it is important to use teachers' objectives and account for these in the selection of testing instruments. Popham, Stake, Edwards, and Colwell emphasize this point and the Music Achievement Tests satisfy not only this requirement as well as being nationally normed in case the teachers desired comparisons between their students and others in the nation, but also the reliability requirements of Marks and Martin. Therefore, it appeared sensible to select an objective criterion of music teacher success and use appropriate treatment techniques even though the literature was antipodal.

The next serious question to search for answers in the literature related to variables other than teacher effects and previous experiences which could affect the gain accountability criterion. Because O'Connor's technique was employed, it was necessary to determine if certain variables such as (1) interval between pre and posttesting, (2) number of class meeting times per week, (3) number of minutes per session for each music class, (4) the socio-economic nature of the school and children, and (5) the time of initial pre-testing could affect the computed residual gain scores and, therefore, should be included as covariates in the equations to derive class residual gain scores. No studies in music education were found which used the residual gain technique as a criterion for music teacher effectiveness. In fact, no studies in music education used pupil growth measurements as teacher accountability criteria.

Hill (1970) investigated the musical achievement of culturally deprived children and culturally advantaged children at the elementary school level using the Primary Music Skills Test which he developed from The Iowa Tests of Music Literacy by Edwin Gordon. Reliabilities for the Primary Music Skills Test were .85 to .98 for test-retest of 24 hours and from .77 to .92 for splithalf estimates using Spearman Brown corrections. The validity of the test was estimated by pooling three judges' ratings of subjects' song responses and computing biserial correlation coefficients to total earned test scores. Resulting coefficients were significantly high with t-test values all exceeding table probabilities at the .001 level (Hill 1970:105).

The sample for his study included 757 children in kindergarten, first, fourth, fifth, and sixth grades from three large midwest cities who came from deprived and advantaged environments. All the deprived children came from homes where income levels did not exceed $3,000.00 per year and where the parent was on some kind of relief roll. Over 50 percent were from broken homes. The advantaged children's family incomes ranged from $7,000.00 to $25,000.00 per year. Only 10 percent of these children were from broken homes. Based on the data analyses through ANCOVA techniques, Hill concluded:

> 1. There was a significant difference between the mean scores on the Primary Music Skills Test of children in the kindergarten and first grade who come from deprived and advantaged environments. These mean differences were all in favor of the latter group.
>
> 2. There was a significant difference between the mean scores on the Iowa Tests of Music Literacy of children in the fourth, fifth, and sixth grades who come from deprived and advantaged environments. These mean differences were also all in favor of the latter group (Hill 1970:121).

Furthermore, he found:

> 1. There was no differential effect on the scores of either test across grade and environmental lines which may be associated with sex. In addition to this fact, there were no interaction effects of sex with either environment or grade, indicating uniformity of performance of boys and girls within and across all categories.
>
> 2. There was a significant effect on the results achieved on the Iowa Tests of Music Literacy Melodic and Rhythmic subtests, when the Iowa Test of Basic Skills scores were held constant for children in the fifth and sixth grades from deprived and advantaged environments. F values between environmental groups were reduced to levels considered not significantly beyond chance. However, levels of significant difference between environmental groups remained for the Harmonic subtest and Composite Test Scores.
>
> 3. Mean scores on the Primary Music Skills Test and the Iowa Tests of Music Literacy can be considered parallel across their prospective grade levels between deprived and advantaged environmental categories. Examination of the means on the Iowa Tests of Music Literacy did indicate a slight but steady increase in the distance between each

> respective environmental category through grade levels four, five, and six, but not large enough to be statistically significant (Hill 1970:121).

Hill concluded that there were established differences in the musical achievement test performances of deprived and advantaged children as judged by the criteria but to what degree these differences may be attributable to environmental factors or menable to remedial procedures remains unanswered.

In his discussion of the third year conclusions of his five year study, Gordon (1972) states:

> At the completion of the first year, it was found that regardless of aptitude level, students who attend culturally-heterogeneous schools were able to perform significantly better on their instruments than students who attend culturally-disadvantaged schools. After two years of instruction in instrumental music, no significant differences were found between the etude performances of the two cultural groups, although culturally-heterogeneous students overall appeared to continue to maintain superior ability in this endeavor. However, by the end of three years of instrumental training, the observed mean difference favored the combined etude performance of the high aptitude students who attend culturally-disadvantaged schools. Nevertheless, the combined etude mean favored the low aptitude students who attend culturally-heterogeneous schools. Therefore, considering the precision of the validity criteria, it appears from the results of only three years of the study at least that students with high musical aptitude who attend culturally-disadvantaged schools are capable of attaining standards in instrumental music similar to those observed for high aptitude students who attend culturally-heterogeneous schools (Gordon 1972:56).

In 1968, Gordon administered his Musical Aptitude Profile to all fifth and sixth grade students in four culturally disadvantaged elementary schools and three culturally heterogeneous elementary schools in Des Moines, Iowa. Every student, regardless of his MAP scores, who volunteered to take instrumental music lessons and to participate in band activities for a five year period was loaned a relatively new musical instrument provided by the National Association of Band Instrument Manufacturers and donated to the University of Iowa. The 230 available instruments were divided equally between the cultural groups.

The criteria used for evaluating students' musical achievement after three years of instruction consisted of ratings of their tape recorded instrumental etude performances and their scores on the Iowa Tests of Musical Literacy, Level 3. Levels 1 and 2 were administered after the first and second years of instruction.

While one may criticize Gordon for using his own tests to validate his other tests and to not use a published performance examination such as the Watkins-Farnum Performance Scales, this current research is welcomed and needed. There have been very few longitudinal projects in music education. After the first year of his study, significant cultural differences were visible. But, by the end of the third year, those students who originally had more aptitude and attended a culturally disadvantaged school did the best - produced the best scores on the criterion instruments. There may be a number of reasons for this finding, one being lack of proper criterion testing instruments. But, even assuming that the ITML is a valid and reliable instrument, which the author so reports, the significance of this study is that students in early high school (secondary school) will vary in music achievement, and one of the possible factors affecting this variance may be their socio-economic and environmental backgrounds; a variable over which researchers have little control other than to treat the data. Hill, too, found significant cultural and social differences but at the elementary school level.

These two research investigations demonstrated that cultural factors could affect music achievement test scores; in other words, teacher accountability in terms of pupil growth could be affected by factors unrelated to the teacher per se such as environmental conditions of the students' homes. This factor was considered to be important in the present study so that schools comprising socially advantaged, socially disadvantaged, or an even mixture of both groups were accounted for in the statistical treatment of the gain score data. This was mandatory in order to provide each participating music educator with the opportunity to produce gain in his students since this was one construct of music teaching success.

Another factor implied in the consideration of pupil gain scores is just what can and cannot be controlled? Furthermore, will music teachers at the secondary school level have as great a potential for producing growth in their students than teachers teaching at the elementary or middle school levels? Will a student participating in a music performing organization have the opportunity to grow as much as a student participating in a general music class, history, theory, or music appreciation class?

These questions led to the research of Folstrom (1967) who compared the music achievement of high school choral performances organizations with students not participating in music classes. Using the Aliferis Music Achievement Test: College Entrance Level and ANCOVA techniques, he found significant differences in favor of the choral groups. In others words, students in these choral organizations learned some specific musical behaviors which could be measured. However, one interesting finding was that several of the control groups' mean scores also differed significantly which indicates that some musical growth occurred for the control group who were without benefit of formal school musical instruction. As Sidnell emphasizes in his critique of Folstrom's dissertation, significant growth was often less than 1 point on each subtest (this significance being due to the large numbers of students participating in the study). The question of statistical versus practical significance must be answered. How does a music teacher teach for a less than one point gain in music class mean scores (Sidnell 1971:63-64)?

The results of Folstrom's research led this investigator to conclude that music educators teaching at the high school level and participating in this study would be well advised to use younger age groups of music pupils for the pupil gain criterion; in other words, they should not use their best performing groups even though this group could be expected to show music achievement and growth. For example, a music educator teaching band in the usual four year high

BEST COPY AVAILABLE

school was encouraged to use his ninth or tenth grade band for the gain score criterion.

The factors referring to number of class meeting times and minutes per session have not been investigated in the literature in any rigorous fashion. There is considerable empirical evidence to support the general conclusion that the more times per week with less minutes per session approach is better than less times per week and more minutes per session. For example, classes meeting for 150 minutes per week would be advised to have three sessions of 50 minutes each rather than two sessions of 75 minutes each.

There are no music research studies in the literature that have investigated the differences in gain attributable to different pretesting times. Empirical evidence suggests that pretesting in the fall is superior to pretesting in winter because of the amount of time the students have had to learn and therefore, affect that pretest score. With this in mind, the design included a check for testing the differences in gain attributable to pretesting time differences. A sixteen week interval for instruction was selected to approximate one semester's instruction time; however, due to human nature, it could be determined that interval would be a factor in this study and could be statistically controlled.

A review of the literature showed that while gain scores were questionable for use as teacher accountability criteria, a considerable amount of effort had been undertaken. No research study could be found which used gain scores as a teacher accountability criterion in music education. Therefore, it was decided that by using the most current thinking as exemplified by Stake, Edwards, Colwell, Lord, O'Connor, and Marks and Martin, a residual gain score could be formulated which would be useful to this investigation. The writer was well aware that this pupil gain criterion was extremely limited in scope, but that it might aid in ascertaining the usefulness of gain scores in music education research and evaluation. Considering the recent moves by state legislatures to mandate state accountability laws, one can hypothesize that if music educators are not included in the evaluative schemes, little money will be forthcoming for music education programs. Since many mandated programs include pupil measurement, it was felt important to attempt a reduced version of the model in this research project. The Music Achievement Test I-IV and the residual gain technique with the aforementioned variables as covariates satisfied the requirements of the best minds currently involved in this area.

Ratings of Teaching Success

A number of rating instruments have been developed for use in the systematic observation and evaluation of teachers and classroom activities. The most famous include Withall's (1949) Measure of Classroom Climate, Ryans' Classroom Observation Record (1960), Flanders' Interaction Analysis, Medley and Mitzel's Observation Schedule and Record (OScAR 1958), and the categories developed by B. O. Smith and others for analyzing the verbal behaviors of students and teachers into ventures. These scales facilitate systematic observation by using categories that are well defined and relevant to the problem under investigation. Many kinds of rating scales are less systematic and require a higher degree of inference on the part of the observer in order to correctly categorize teachers' and pupils' behaviors. It is important to recognize that ratings, no matter how systematic and specific, are still opinions of the rater about the ratee or about the process of teaching and the environmental conditions.

Rating advocates criticize pupil gain measures for allowing too much unaccounted for variance to enter into the gain score picture. The products of pupil behaviors as indicated by gain scores are, in essence, not toally attributable to the teacher because of other extenuating circumstances: environmental, sociological, intellectual, and additional uncontrollable variables. Because there is no assurance that the teacher's behavior is the true cause of the students' behaviors, Ryans argues that evaluation should concentrate on the behaviors observable in the classroom. This is in line with Flanders who prefers to describe what goes on in the classroom. In this sense, Stake and Flanders agree as to writing evaluation descriptors which aid in the decision making process. Flanders states:

> First, the heart of the matter lies in what the teacher does that influences the educational development of his pupils. Second, of all teacher activities, the most salient are the direct person-to-person contacts and the more indirect teacher-to-class contacts. Third, we start by accounting for events that do, in fact, occur, not what the teacher thinks is happening or what teachers "ought" to be doing. Fourth, the most important concepts are those which are descriptive of the interactive contacts. Other concepts in a field of education simply become less important; these are set aside to be learned and used only after the higher priorities have been satisfied. Fifth, the search is for the fewest number of ideas necessary to help a person develop and control his teaching behavior (Flanders 1970:3).

Flanders' concept of classroom interaction refers to the chain of events which occur, each occupying only a small segment of time. These sequences or chains of events add up to a pattern that can be identified, occurs frequently enough to be of interest, and can be given a label or name since this often facilitates thinking. He says:

> Ever since 1900 researchers have attempted to evaluate teaching performance without too much success . . . Since 1952 the search to find teaching acts which are significiantly and consistently correlated with positive pupil attributes and content achievement adjusted for initial ability has been much more successful . . . The progress in this area was possible, for the most part, because of the development of systems for analyzing classroom events (Flanders 1970:6).

Flanders feels that when teacher education focuses more sharply on the case and nurture of teaching behavior, systems of interaction analysis could become the foundation of a program for preparing teachers. The ultimate criterion of success or failure can be found in the classroom performance of beginning teachers. "By assuming that classroom interaction is a series of events and that teaching behavior consists of acts, or patterns of acts, embedded in the chain of classroom events, then a first step is to break down the patterns of teaching behavior into teachable skills" (Flanders 1970:8). The logic then would be to break these patterns down for teaching purposes into a sequence of events from simple to complex. So, the ultimate goal is for preteachers to discover, explore and investigate for themselves relationships between teaching behavior and classroom interaction.

Flanders' system of interaction analysis concentrates on the verbal interactions occurring in the classroom. Codings are made every three seconds, and a pattern for the class period evolves.

Ryans' *Classroom Observation Record* identifies three dimensions of classroom behavior: (X) includes the friendly, understanding versus aloof teacher, (Y) refers to the responsible, businesslike, systematic versus the unplanned, evading, and slipshod teacher, and (Z) refers to the stimulating, imaginative, surgent versus the full routine teacher. It might be assumed that a good teacher will lean towards the positive ends of the X, Y, and Z scales. Whether these three scales are discrete is open to investigation.

The Smith and others' *Strategies For Teaching Scale* classifies ventures into the following areas: causal, conceptual, evaluative, informatory, interpretive, procedural, reason, rule, and system. Each venture includes subpoints of classification as well as levels which are similar to the Bloom and others' *Taxonomies*.

These three rating scales are the most complex and probably the best available for use that are standardized and readily available. Nolin (1969) adapted the Flanders' scale for use with junior high school music classes and Erbes (1972) adapted it for use with music performing classes. The only difficulty with this approach is that nonverbal communication is not considered. So often and especially in music education the wink, smile, frown, nod, and so forth are effectively used to relay positive and/or negative reinforcement.

Furthermore, the nature of the art requires that students in music classes respond musically, not verbally. If a music educator desires a pupil to make a musical response, the pupil might play or sing, clap or tap, move or jump in indicating the musical behavior. Verbalization is used but not to the extent that it is used to learn concepts in the science and mathematical areas of the curriculum. It is quite possible that for this reason Nolin found very few statistically significant differences between general music teachers rated most and least effective (Rainbow 1971:45).

For the above reasons, it was decided not to use one of these well researched rating scales in this present study. Furthermore, the Flanders' Interaction Analysis Technique requires skilled raters: persons who have had some training in interaction analysis. This writer could not have visited every participating music educator's classroom in order to obtain data. A rating instrument that could be used by inexperienced raters, that was relatively inexpensive, and that could be used by various raters in different categories was desired. The time element was important, too, so that the rating instrument would require no more than 10 to 15 minutes completion time.

As reported in Chapter One, very few research studies in music education have used pupil evaluations of teachers as teaching effectiveness criteria. These that did will be discussed below. For this study, it was necessary to use an instrument that was (1) specifically designed for secondary school pupil evaluations of teachers and (2) could be modified for gathering rating data in other categories. The only instrument that is published, normed, and standardized in this area is the Illinois Teacher Evaluation Questionnaire (ITEQ), and was, therefore, selected for use in this investigation. An alternative would have been for the writer to develop and pilot a rating instrument. This was not accepted as the most viable procedure since a published instrument was available.

There is little reason to pursue in this chapter the pro and con arguments for rating and pupil gain evaluations as teacher accountability criteria.

The literature is quite diverse with respect to this area. Therefore, the writer needed to formulate certain philosophical value judgments evolving round the best possible approaches for sampling the criterion domain while being fair and considerate to the participants. For this reason, it was decided to use not only rating and gain evaluations, but to also treat them as a subproblem for investigation of the relationships and interrelationships among these two areas or domains.

Pupil evaluations of teachers have been investigated in some studies: Erickson (1954), Albert (1941), Anderson (1954), Watkins (1958), and Lutz (1963) as well as those reported by Silberman (1971) which demonstrated that pupil ratings of student teachers have little relationships to other ratings such as supervisors. The studies excluding those reported by Silberman reflect the low correlations found among various categories of ratings, especially those comparing pupil to teacher evaluations. Because relatively little research has been directed at pupil evaluations of teachers in terms of ratings, it was considered that pupil ratings might portray an additional dimension in the criterion domain, an indication of how the teachers' were accepted by their pupils. The reasoning follows that if there is a tendency for pupils to accept their teachers there might be a relationship between pupil learning as measured by gain scores and pupil evaluations as measured by ratings. While the literature did not assist this study much in this area, it was considered worth investigating and pupil evaluations were incorporated into the design.

Building principal and supervisor evaluations in terms of ratings have been the most common teaching effectiveness criterion found in the literature (Guilford 1954:263). Earlier studies with elementary classroom teachers also used the school district's superintendent. In the field of music education, many studies have used the music supervisor or administrator as well as sometimes offering a composite supervisor-administrator criterion. As has been mentioned previously, research in the criterion domain has generally concluded that principal, supervisor, and administrator ratings are reliable and consistent; yet, they do not relate significantly to pupil gain or pupil rating measures. This investigation decided to investigate ratings per se and not to use a composite administrator category including the supervisor of music's and building principal's ratings. Because it could be anticipated that a portion of the participating music educators would not have a school district's music supervisor, a missing data supervisor rating was predicted for them. Most research studies have assumed that a building principal can evaluate his teaching staff; yet, questions raised by Funk (1973) which show significant differences in personality profiles between groups of music educators and building principals lend credence to the use of separate rating categories.

Peer evaluations have not been investigated in any rigorous fashion in the elementary and secondary schools. Lutz (1963) used one peer nomination with his study in high school instrumental music education. However, at the college level, groups of professors may meet and advise the administrative staff on merit pay and promotions for peer professors. However, in the lower grades, peer teachers generally do not rate other teachers unless they are department heads. Therefore, it was speculative as to whether peer music educators and other teachers in the participants' building could rate them even though they may have never seen the music educator teach music. Since little guidance was found in the literature, it was decided to obtain 5 peer evaluations per teacher and construct a mean peer rating score.

Self ratings have not been investigated other than by Erickson (1954). Many self inventories have been studied but these do not refer to the teacher's specific abilities as a teacher or his perceptions about himself as a teacher. Borkowski (1970) used a self rating of major instrument music playing ability as an experimental variable and found these to be somewhat negatively skewed.

Because ratings were the only criteria used for the music student teacher cross validation group, the same five categories were used. Research by Anderson (1965), Morgan (1946), Turrentine (1962), and Watkins (1958) indicated that in music, ratings of student teaching effectiveness could be used even though Morgan found rankings to be superior. Because seven institutions' music student teachers were used in this study, the rank order procedure would have been too difficult to obtain and use. Pupil, cooperating teacher, supervising teacher, music student teacher peer, and self ratings comprised the rating criterion for this group. The main difference between the two groups used in this study was that for the peer evaluation category, music student teachers rated each other while the peer category for the music educator sample included other peer music educators as well as teachers of different specialities who knew the participant well.

The literature is well established as far as rating criteria; yet, the interrelationships among the ratings appear to be in need of validation in music education. For this reason, relationships and interrelationships among the criterion variables were investigated and rating categories could not be accepted as equal in either dimension or domain. Appropriate statistical techniques were used in order to investigate these relationships and provide proper weightings for deriving composite ratings for each group under investigation.

To reiterate, ratings are subjective by their nature - they are the rater's opinion of the ratee's performance or success. Much enters into these evaluations that cannot be controlled to any degree of uniformity or accuracy. Yet, as suggested in the literature that by using the same rating form for all the ratings, some error variance should be avoided. Furthermore, by expanding the rating criterion to include 5 ratings per teacher and music student teacher, it was felt that this approach would be most fair to the participants.

From the above two main dimensions - pupils residual gain scores (G) and ratings of teaching success (R) - four criterion variables were formulated for the music educator sample: (1) pupil residual gain (G), (2) ratings of teaching success (R), (3) a composite criterion of gain plus rating (G+R), and (4) a discrepancy criterion of gain minus rating (G-R). One criterion would have been researchably indefensible according to the literature because (1) studies by Barr and others have shown little relationships among the criteria, (2) fairness to participants in terms of sampling the criterion domain was considered important, (3) the major criterion variables found important in other studies should be incorporated into the design of this present investigation, and (4) the means by which the criterion domain would be treated in order to arrive at the derived variables should be fair, logical, sophisticated, and statistically sound. Factor analytic techniques appeared to be the best for formulating a composite rating (R) criterion as indicated in the literature. Therefore, these techniques would be used in this present study.

This present study was not experimental but used experimental techniques to a certain extent. Any descriptive study of this nature must develop a theoretical construct of teacher accountability. The criteria employed in this study aided the formulation of the four constructs derived - Gain, Rating, Gain plus Rating, and Gain minus Rating - so that predictive relationships among them and the experimental psychometric variables could be investigated. It was accepted at the outset that certain of these criteria might be highly questionable in light of present day thinking; yet, with current trends to product and process measurements, it was deemed essential that both, to some extent, be included in this research design. To close, the most difficult and time consuming aspects of this investigation were related to the music teaching and music student teaching success criteria. For this reason, they were chosen with great care and treated with the most reliable and valid statistical means found appropriate in the literature and were available to the writer and his committee. As will be shown in Chapter Three, a considerable amount of time and effort was expended in order to assure that the criterion demands of this study were satisfied. No one investigation can include every portion of the teaching act; but, within the scope of this study an attempt was made to be as thorough and comprehensive as was humanly possible.

SECTION THREE: THE RELATIONSHIPS OF PERSONALITY TO MUSIC TEACHING SUCCESS

Section One of this chapter highlighted the pertinent research studies in the literature that indicated that personality variables could be used in a design to predict teaching and music teaching success to some degree of usefulness. This section concentrates on those studies most pertinent to the design of this present investigation and emphasizes the techniques used to investigate the relationships and interrelationships among personality variables and success criteria.

There are persons today who oppose prediction studies on the thesis that once one can predict he should implement selecti[illegible] retention programs. These writers would prefer researchers to investigate the relationships and interrelationships among variables to success criteria and report the results without formulating predictive claims. This is in fact what most prediction studies have done in the past, and these views are acknowledged in this study. Before one can predict, he must first investigate the relationships, interrelationships, and interactions among the independent and dependent variables and offer reasons for any significant findings. Prediction implies fact - the fact that with some battery of tests, one can include or exclude invidivuals according to their battery test scores. These scores form a profile and these profiles are then compared to profiles of successful persons in the occupation for final predictive judgment. In industry, where highly discrete tasks are readily discernible as success criteria, predictions are valid and useful.

BEST COPY AVAILABLE

For example, Rawls and Rawls (1968) were able to significantly differentiate successful and less successful executives using just personality (EPPS) and personal history data. Criteria included: (1) salary level, (2) company job title, (3) job number as listed in the Hayes Salary Survey, and (4) appraisal performance ratings from at least 3 former supervisors (Rawl and Rawls 1968: 1033).

In education, predictions can be made but only in light of certain, less specific, success criteria. Evaluating executives and evaluating teachers are completely different processes and require unique models. It is in this light that because errors can be made, one should consider a prediction study or predictive results as aids along with a host of other tools for counseling and advising rather than selecting and excluding. It was this fundamental philosophical position that the writer cautiously offers the following studies in the literature and interprets them in light of his selected design for this research investigation.

The area of personality is complex. One must consider separately those variables that refer to basic personality traits and behaviors from those of the dynamic domain which refer to interests, drives, values, attitudes, feelings, and sentiments. Most research studies have found some significant personality trait correlates but very few have investigated the dynamic domain in depth. The reason for this is because the personality trait instruments have been developed, tested, and validated over a considerable period of time. The tests most commonly used in the literature have been the Minnesota Multiphasic Personality Inventory, the Guilford-Zimmerman Temperament Survey, the Edwards Personal Preference Scales, the Sixteen Personality Factor Questionnaire, the Thurstone Temperament Schedule, the California Psychological Inventory, and other tests which measure specific traits such as dogmatism, surgency, intelligence, dominance, agressiveness, masculinity, and so forth. Interest inventories are available but have lower reliabilities than the trait tests. These include the Minnesota Teacher Attitude Inventory, the Strong Vocational Interest Blanks, and the Kuder Preference Tests to mention but a few. In general, these tests measure surface attitudes, are easily faked, and do not measure the more ingrained attitudes and motives such as values and morés (Getzels and Jackson 1953:520).

Because of the difficulty in personality and motivational measurement as well as in the area of defining teaching success, little consensus has been found in the literature. Traits that have been empirically derived appear to correlate to some rating criteria while showing little relationships to gain criteria. These findings may be more affected by the criteria than by the personality variables.

Very few studies have investigated motivational variables. The Motivation Analysis Test by Cattell and others was selected for use in this research project because it attempts to delve deeper into the realms of drives (ergs) and sentiments, is not easily faked, and is more fun for participants to take than the commonly used attitude instruments. Only two previous investigations have used the MAT test probably due to its cost and difficulty [complexity] of scoring procedures. Handscoring is laborious, and machine scoring expensive. However, MAT is the only test which measures 10 motivational variables in each of two categories: unintegrated which reflects those dimensions that are internal and unconscious and integrated which reflects dimensions which are conscious and therefore subjective and social in nature.

The only study using both 16PF and MAT variables was conducted by Holland Beazley (1970) with students majoring in forestry at the University of Illinois and Southern Illinois University during the period 1962-69. Answers to the following questions were sought:

1. What sort of individual is the forestry student?

> 2. Are the personalities and motivations of forestry students different from those of other college students?
>
> 3. If there are differences, what are the implications for better professional performance and/or curricula changes which might lead to better performance?
>
> 4. How can the self-selection of students in forestry be improved (Holland and Beazley 1970:421).

Their concern centered round the newly created problem of professionally trained foresters to successfully manage forest and related land resources within rapidly changing patterns of demands for forest products. The motivations and traits of persons with managerial abilities may differ from those of the general forestry student.

Using t-tests across the 16 variables of the 16PF and the 20 MAT variables and comparing these to the average male college student scores as found in the norms in the Handbooks, they concluded the following:

> In general these findings, in regard to the primary personality characteristics (16PF), indicates that the forestry student, in comparison with the reference population average, is more reserved in manner, more intelligent, more emotionally stable, less demanding, more enthusiastic, more conscientious, more toughminded, more forthright, somewhat less assured, more conservative, and more self-reliant (Holland and Beazley 1970:421).
>
> A general interpretation of these motivational characteristics suggests that the forestry student, compared to the average male college student, is more concerned with material activities, expresses hostilities naturally, is willing to meet the world "on its own terms," and tends toward independence and autonomy. Furthermore, forestry students appear to lack an egocentric life style, have a low need for career status and economic competition, to be low in anxiety, and generally satisfied and contented (Holland and Beazley 1970:422).

The profile does not suggest a person well suited to attainment of leadership in the management of either public or private forest resources where people, politics, and economics are heavily involved. Therefore, the complaint that too few foresters reach the top in major forests products companies appears to be well founded. Foresters make excellent technicians and researchers. "There is little question that the successful development of the U. S. Forest Service, for example, as an effective, decentralized public forest resources administering agency has been due in large measure to the kind of people attracted" (Holland and Beazley 1970:422). They concluded by stating that forestry schools will need several kinds of action for preparing foresters, managers, technical consultants, and so forth. They suggest that a program of personality and motivational testing be continued so as to counsel and advise prospective forestry students and to distinguish between the students who are technically and materially oriented from those interested in management.

While one may question the use of t-tests across each variable with little consideration for the interrelationships of the variables, Holland and Beazley have shown that personality and motivational variables do differentiate forestry students from the general college population. This was also found in the literature to some extent for music students, but whether these variables separate among predicted success levels is the unanswered crucial question.

Cattell, Barton, and Dielman (1972) used the Culture Faire IQ Test (CFIQ), the High School Personality Questionnaire (HSPQ), and the School Motivation Analysis Test (SMAT) to predict four standardized achievement test scores (ETS) in the areas of mathematics, science, social studies, and reading. There was a 3 month interval between the prediction and the criterion testing. Subjects included 169 sixth graders and 142 seventh graders attending a middle school of upper middle and lower class pupils.

They found that personality, motivation, and ability variables independently contribute to the variance found in achievement scores even though the relative importance of these three types of variables varies according to the area of achievement considered. "Because the HSPQ contains a measure of intelligence (Factor B), in most areas a high level of prediction can be obtained by using the variables of SMAT and HSPQ together without the CFIQ measure" (Cattell and others 1972:43).

An interesting finding was that with these grade levels, the unintegrated scores were virtually useless in the prediction of achievement and appeared to act as suppressor variables. They suggested that counselors and educators should still use both in order to shed light on the relationship between Unintegrated Motivation Scores and achievement. They obtained a multiple R of .71 which, in the area of mathematics, the variance accounted for reached 61 percent in the sixth grade and 69 percent in the seventh grade. It is interesting that they found such a high multiple R considering that the SMAT is less reliable and has not been used very much. SMAT is a version of MAT designed for secondary school use. No studies have been found that used the MAT test with teachers so there was no related literature to guide this present research investigation.

Differences Among Teaching Speciality Groups

As reported in Chapter I, Bodenstab (1968) found psychometric differences among 11 teaching interest areas including music and among the grade level groups of elementary, special, and secondary. Using a MANOVA design, he concluded that the teaching profession is psychologically heterogeneous with respect to interest areas (Bodenstab 1968:319).

Studies by Wink (1967) and Funk (1973) bear directly on the problem of heterogeneity. Wink found that his group of music student teachers differed significantly from the national norms: higher in self concept, deference, and order while lower in affliation, intraception, and anxiety (Wink 1970:241).

In a study of 55 high school music teachers and 55 administrators selected from 34 southern Arizona high schools, Funk (1973) investigated differences in personality traits and attitudes about music education between the two groups. He used the 16PF-Form A to gather basic personality trait information and developed a 69 statement opinionnaire which afforded respondents a choice among 6 reactions to controversial statements. Using t-tests for the personality variables and chi square for the opinionnaire variables for checking differences between the two groups, Funk found 16 of the 23 personality factors (he used primary and secondary scoring procedures) beyond the .05 level between the group means of music teachers and administrators.

> The salient personality factors characterizing the group of administrators were self-discipline (self-control) and will power, conscientiousness, leadership, social boldness, emotional stability, warmth, imaginativeness, social polish, and sensitivity. They appeared to be relaxed, self-assured, and confident, trusting, well-adjusted and free of anxiety, apprehensiveness or feelings of guilt. They were conservative; they lacked objectivity and tended to react on the basis of feeling rather than thought.
>
> The music teachers were characterized as being outstanding in the areas of creativity, social polish, sensitivity, conscientiousness, self-sufficiency (independence), neuroticism, imaginativeness, tenseness (feeling "driven"), intelligence, self-discipline (self-control) and will power and social boldness. They are conservative, trusting, and introverted. They lacked objectivity and were seriousminded rather than gay and carefree.
>
> Several apparent contrasts between the two groups should be considered especially important because these differences, if present to a marked degree between an individual principal and music teacher, might make for difficult relations between them: (1) The administrators appeared to be outstandingly warm and easygoing, while the music teachers were more cool, reserved and detached; (2) The administrators were relaxed and emotionally stable, while the music teachers tended to be tense and easily upset; (3) The music teachers were much more inclined toward independence than were the administrators. The music teachers appeared disinclined to follow and preferred their own decisions. They appeared to be self-sufficient, in contrast to the more group-oriented administrators (Funk 1973:145-147).

When group responses were compared item by item on the opinionnaire using the chi square statistic, differences were found to exist beyond the .05 level for 32 of the 69 items. Of the 12 statements related to pay and working conditions, significant differences were found for 6. Those statements dealing with scheduling, auditions, and rehearsals, responses were significant for 6 of the 12 total. In the area of philosophy and course content, significant differences were found for 3 of 11 statements. In the fourth area of performance, policies, and public relations, 5 out of 12 elicited significantly different reactions. The fifth area included six statements about problems of budget, equipment, physical plant, and transportation of which 3 were statistically significant. The sixth area of controversy, that of direct person to person relations between music teachers and administrators, elicited 7 of 19 significantly different responses. The seventh area included statements ascribing a typical personality to administrators and music teachers and only 2 of the 6 items were significant. Therefore, Funk found considerable differences in opinions between groups of music educators and administrators across these areas (Funk 1973:126-136).

> The remarkable uniformity and consistency of results obtained in measuring personality factors, values and attitudes of teachers in all areas of the creative and performing arts, lend weight to the conviction that a selected group of evaluative tests in these areas might prove to be a valuable predictor of success in the creative arts field (Funk 1973:159).

Funk's study was neither predictive nor differentiative of personality and opinionnaire differences among goodness levels of administrators and music educators. He simply confirmed previous findings that personality profiles differ among teacher speciality groups.

Furthermore, this study could be misleading. The fact that statistically significant variables were found to differentiate the groups does not mean that the results will hold in replicative studies nor would these variables be useful in applying regression analysis techniques as implied by his conclusion relating to predicting success in the creative arts field.

The third in a series of studies of music therapists who work in hospitals and rehabilitation settings concentrated on the "Personality Traits of Music Therapists" (Shatin and others 1968). The 16PF was used to contrast mean scores between the music therapist group and a musician performer group.

Two traits significantly differentiated the groups: sober-lively and expedient-conscientious. A third, trusting-suspicious, showed a tendency toward such differentiation. These findings suggest that the music therapists are more prudent, more conscientious, and tend to be more trusting than their musician performer counterparts. These differences lie within the central area of the profile and cannot be accepted as definitive according to the researchers.

When comparing the music therapist group to the general population, the sten scores suggest that the music therapist group is substantially more intelligent and tenderminded, somewhat more conscientious and self sufficient or resourceful and more trusting and placid than the general adult population. A similar comparison of the musician performer group suggests that this group is substantially more intelligent, more happy go lucky or lively, more venturesome or spontaneous, more tenderminded or sensitive, and more self sufficient or resourceful. These findings suggest that a similar profile exists for groups of musicians and that personality variables may not differentiate to any substantial degree music therapists, music educators, and performing musicians. This study supports the others found in the literature that significant differences are to be found when comparing musician groups to other professional and occupational groups in the general population. It also hints at possible profile differences within groups of musicians.

Student Teachers in Music.

Morgan (1946) attempted to predict student teaching success for senior music students at the University of Minnesota. Using a battery of tests, she investigated 22 traits and abilities (see Section One of this chapter). The following 14 were found to be significantly related to music student teaching success:

1. High school centile rank
2. American Council Psychological Examination Score
3. Cooperative English Examination Score
4. MPI - Morale
5. MPI - Social adjustment
6. MPI - Family relations
7. MPI - Emotionality
8. MPI - Economic conservation
9. Seashore - Pitch
10. Seashore - Loudness
11. Seashore - Time

12. Seashore - Timbre
13. Seashore - Rhythm
14. Seashore - Tonal memory

She found that the ranking procedure was more valid than rating evaluations. Furthermore, students with above median scores on the significant variables as freshmen maintained their superiority throughout the four years of the music education curriculum while students with below median profiles tended to stay below the median throughout the program.

Morgan's study, to a certain extent, was used as a model for this investigation. First, she was very thorough and scholarly in reporting data. Second, she used the best and most sophisticated treatments of the variables for her analyses. For example, rather than just accept a composite rating score a priori, she computed interjudge reliability and then used a partial correlational approach to derive a residual rating score criterion for each participating music student teacher. Both for rankings and ratings, she stratified the music student teachers into 4 criterion groups. The instrument she developed was thoroughly piloted and tested before use in the study. Significant differences in profiles were found between successful and less successful groups of music student teachers. This led this investigator to decide that (1) personality could be measured, and (2) personality variables alone should comprise the experimental variables.

Duda (1961) used the MMPI along with other experiential and background information to predict student teaching success at the University of Illinois with the Ryans Classroom Observation Record as the criterion. Criterion variables included Pattern X which refers to friendly, understanding versus aloof teaching behavior, Pattern Y which refers to the responsible, businesslike, systematic versus the unplanned, evading, and slipshod teaching behaviors, and Pattern Z which refers to stimulating, imaginative, surgent versus dull routine teaching behavior. He found that he could predict to some degree of accuracy Patterns X and Z for music student teachers.

Much of the student teaching experience involves working with performing musical organizations which are primarily teacher directed and conducted. The Record was designed for use with regular classrooms and could, therefore, be considered inappropriate for most music teaching situations with the possible exceptions of general music, history and theory, and literature and appreciation classes in which the prevailing classroom atmosphere might be similar to other school classrooms; but, its use with performing musical organizations such as bands, orchestras, and choirs must be questioned. Yet, Duda is to be commended for desiring a criterion instrument which was well tested and standardized in the field and was reliable and valid.

Evidence exists that classroom behavior per se is not the prevailing problem in teacher weaknesses nor reasons for teachers leaving the profession (Brown 1955). These teacher weaknesses are not recorded by Ryans' Record, yet, they might be inferred. Therefore, criteria for evaluating music student teaching success should relate to the act of student teaching in music.

Foregoing the criterion problem which all research studies must acknowledge to some extent, it appears that Duda really did not predict student teaching success in music per se. What he really did was to investigate some relationships using regression techniques; and, while these are the main techniques for prediction studies, the ascertaining of significant multiple regression equations is not enough. These equations must be tested on a different sample of music student teachers. The usual technique is to divide the sample in half and formulate equations for each half and cross validating between groups. Another procedure is to formulate the multiple regression equation on one group and use the weights derived from the first group to make predictions for the second group which is usually smaller in size. Simple correlations between the predicted and actual scores will provide cross validation information. Therefore, it has not been demonstrated whether Duda's regression equations would really predict success for another group of music

BEST COPY AVAILABLE

student teachers. It is possible that in 1961 the need for cross validation was not emphasized as much as it is today in the scholarly journals.

The significance of Duda's study for this present research investigation is that he was able to predict for certain behaviors as measured by Ryan's Record: primarily Patterns X and Z which refer to friendly, understanding, stimulating, imaginative, and surgent music student teachers. The rationale for these findings could be that underlying these criterion behaviors are basic personality traits or variables. Since he found some MMPI linkage to these patterns, it would have been ideal to see if these really were significant with a music educator group. In other words, if Duda could predict success for music student teachers using a criterion instrument that, in the opinion of this investigator, was invalid for his group; then, it stands to reason that the personality domain was interacting to such an effect as to affect his regression equations. Were the independent variables really independent or was he really showing that the criterion variable patterns were linked to personality patterns? If this were so, this could determine why he found significant equations when his criterion related little to what a music student teacher actually does in the act of music student teaching and in producing musical learnings in children!

Duda is to be commended for researching an area of interest to all music educators. Student teaching is crucial to the predictive potential for music teaching success. The only problem is can one find criterion variables similar for both student teaching and music teaching? Popham with his teaching tasks criteria thought that he could, but even this approach needs further refinement. Because both Morgan and Duda found significant predictions for music student teaching success, it was decided to see if music student teachers could be used as a cross validation group to substantiate predictions, if any were found, for music educators on the theory that first a prediction equation must be generated for showing a personality profile for the successful music educator. If this were possible, then music student teachers who were successful may show a similar profile and could be predicted. Both Duda and Morgan used the MMPI, Morgan the earlier, experimental MPI version. Both found personality variables significantly related to music student teaching success. Both used good statistical data treatment. Therefore, it was decided to use music student teachers but to use as many as possible from different teacher training institutions. By using seven institutions' music student teachers, a breadth of experience and abilities was desired which could reflect the multifarious characteristics of the music educator sample group.

Turrentine (1962) used The Ohio State University Psychological Test college entrance examination board's Scholastic Aptitude Tests - verbal and mathematic, Otis Quick Scoring Mental Ability Tests - Grade 9-16, high school centile rank, sophomore year GPA, music theory GPA, and teacher training courses' GPA for music student teachers at Lawrence University. Like Duda, he used multiple regression techniques without cross validation and found methods courses and high school centile rank related to music student teaching success which, in this case, was the student teaching grade. Turrentine can be questioned for using the music student teaching grade as the success criterion since these tend to be inflated and negatively skewed. Previous research was supported in that grades and previous experiences are related to predicting music student teaching success. However, by using the grade point averages of several courses to predict a music student teaching grade, he may have found simply a relationship that is known to exist. Grades predict future grades! The 80 music student teachers in the sample appeared sufficient for cross validation. It is this writer's opinion that the GPA criterion variable was not sufficient for discriminating among the music student teachers: the other possibility was that Lawrence University has only superior music student teachers.

Anderson (1968) used the 16PF and concluded, like Turrentine, that the best single predictor was the total GPA. None of the other variables, especially those personality trait variables from the 16PF, was significant.

Anderson (1965) used a composite rating of the cooperating and supervising teachers as the criterion of music student teaching success for students at the University of Southern California. As mentioned in Section One of this chapter, he found three significant variables related to music student teaching success: (1) the objectivity factor GZTS, (2) the masculinity factor GZTS, and (3) the symbol production factor of creativity from the Project Potential Tests of Creativity; all at the .05 level. Because he had only 19 music student teachers, he felt that regression analyses would have been undesirable for predictive purposes. However, since he had 13 experimental, 1 criterion, and 19 music student teachers, he probably could have tried to develop some multiple regression analyses rather than using chi square techniques. Certainly, he did not have sufficient numbers to cross validate.

Anderson's use of creativity demonstrates the difficulty of measuring this area objectively and effectively. The Guilford-Zimmerman Interest Inventory was probably too unsophisticated to measure more than superficial interests. The use of the Kwalwasser Music Talent Test must be questioned since there were at least three better, more valid, and more reliable instruments of musicality available: the Wing, Gordon, and Seashore Tests. Anderson does not substantiate his selection of the KMTT.

While Anderson claims prediction in his title, it did not appear in the body of the thesis. His study further substantiates the hypothesis that personality variables are significantly related to music student teaching success. Of the three significant variables, two were personality: objectivity and masculinity. Furthermore, he found few differences in test results between high potential music student teachers and master teachers. The lack of difference in interest might be due to the high degree of selectivity in the sample and the lack of discrimination in the testing instruments.

Anderson is to be commended for investigating the use of published, standardized instruments to music student teaching success. Furthermore, he conducted a pilot study to determine some characteristics of successful music teachers inservice. He did not compare this sample with unsuccessful music teachers nor did these teachers complete all the tests. However, the crucial point that affected this present study was that Anderson attempted to derive certain characteristics from music educators first before studying music student teachers at USC. This plan of validity attack was adopted in the design of this present study. Also, published, standardized instruments for both experimental and criterion variables were used. While one may question some of Anderson's statistical analyses, his study was thorough in concept, well designed, and well written. It assisted this present study because the writer decided to concentrate on the music educators first, to obtain the best possible stratified sample, to develop predictive regression equations for this sample, and then cross validate to music student teachers while using published, standardized tests for data gathering.

Wink (1967) investigated the relationships between the self concept, personality need system, and anxiety level of music student teachers and their achievement in music student teaching. The subjects were 32 music student teachers at The Ohio State University School of Music during the 1966-67 school year. Using a composite rating gathered from a team of college supervisors and the cooperating teachers for the criterion, he found that personality variables and self concept significantly discriminated between successful and unsuccessful groups of music student teachers.

Wink's study was not predictive, but he did find significant relationships. One significant finding was that personality (EPPS) and anxiety (IPAT) did not change from before student teaching to afterwards. He found that a high self concept was related to dependent, conforming, and gregarious music student teachers. The well adjusted students exhibited less anxiety and the high achievers indicated a need for deference (to yield in opinions, judgments, and wishes and having respect for supervisors or cooperating teachers), order, and affliation with a low need for autonomy. These are interesting findings because they show a possible personality profile for a highly rated music student teacher to be conforming to the wishes and desires of supervising and cooperating teachers. Yet, studies with music educators (Funk 1973, Lutz 1963, Barth 1961, and Fosse 1955) have discovered different personality patterns. The significance of Wink's findings indicated to this investigator that: (1) there may be differences in profiles due to the criterion variables between successful music educators and successful music student teachers, (2) if this were so, the music educators should be studied first, and (3) that the better music student teachers may be conforming due to the authority images of the supervising and cooperating teachers. The fact that Wink found a high self concept significantly related to music student teaching success suggested the possibility that further investigations in the motivational areas might produce positive results.

Williams (1958) developed a Student Teaching Inventory which was mailed to cooperating teachers' former student teachers. He attempted to correlate these opinions of the cooperating teachers' effectiveness with ratings given by the University of Illinois supervisors. He showed a positive correlation. Using the MMPI, the IPAT Music Preference Test, and college records to gather data for his investigation, he was able to derive significant multiple correlations for predicting cooperating teacher success. However, the first question that must be answered is how well can former student teachers evaluate their cooperating teachers? If this is valid, does the passage of time interfere? Most persons, after a period of time, tend to remember primarily the good events, forgetting the unpleasant ones. Even if he could predict well, he provided no procedure for matching music student teacher and cooperating teacher.

While he found some significant relationships among success and psychological factors, practical limitations dictate that less able cooperating teachers may have to be used because of travel and availability considerations. While his attempt to refine this selection process was worthwhile, he did not succeed in having such a program implemented at the University of Illinois.

The importance of Williams' findings is that, to a certain extent, cooperating teachers' effectiveness or success can be predicted. This aided this present investigation by further supporting the claim that the emphasis should be directed to the music educators. Yet, it would be very difficult to investigate the cooperating teacher, too, although many cooperating teachers of participating music student teachers also were participating music educators in this present investigation. Few studies in music education have attempted to study the cooperating teacher, student teacher training variables. Many opinions have been offered but objective and thorough investigations still need or await inquiry. Williams' is to be commended for a first effort in this area.

Music Teaching Success

The domain of predicting music teaching success from college experiential data is unclear. While Benner (1962) found student teaching GPA, music methods GPA, and professional methods GPA related to music teaching success and Strub (1957) found excellence in college grades significantly related to music teaching success, Borkowski (1970) found little relationship between the quality of work in undergraduate courses leading to a music education degree and teaching success. Borkowski's study is more thorough and complete. He used care in selecting his criterion variables: (1) pupil performance,

(2) pupil knowledge of music history and theory, and (3) band performance. He used a jury of experts for the first and third criteria and an investigator prepared music achievement test for the second. The significance of this study to the present one was the fact that he used two measures of pupil success as indicators of music teaching effectiveness. Fosse (1965) argues that band performance ratings are also pupil success criteria since it is the band, composed of students, that is being used to indicate the success levels of the band directors. If this is so, Borkowski, then, used all pupil type of evaluations as constructs for music teacher accountability. Furthermore, the fact that Borkowski failed to find college experiences as reflected by grades and self performance ratings on the band directors' major instruments related to music teaching success, the writer decided to not include these variables in this present study but to use pupil evaluations as music teacher accountability criteria. It is possible that the reasons Benner and Strub found significant relationships and Borkowski did not were due to the criteria, ratings in the first two cases, and pupil evaluations in the latter.

Lutz (1963) investigated the professional backgrounds, experiences, and personality characteristics of 103 instrumental music teachers in Kentucky and divided his group into successful (N=75) and unsuccessful (N=28) on the basis of a composite rating derived from their administrator-supervisor, a fellow peer teacher, and a student. He found the following differences for the instrumental music teacher group as a whole when compared to the general population:

1. High school instrumental music teachers were more sensitive and emotionally labile.

2. They experienced physical sickness more frequently.

3. They were less likely to be concerned with social morés and were more likely to commit asocial acts.

4. The high school instrumental music teachers tended to be more self critical and greater worriers.

> 5. They tended to be more moody and hostile but the hostility was directed inward toward themselves rather than outward toward others (Lutz 1963:121-122).

For the successful group of instrumental music teachers, Lutz found that they were less committed to compulsive or neurotic behaviors, were less moody and better able to reconcile internal concerns, were more capable of deep emotional responses, were better able to profit from experience, had a higher degree of emotional morale and were better able to manifest normal optimism with regard to the future, strived for a higher degree of social approval, were less hostile, worried less and were more flexible in meeting day to day situations, had broader interests, were less subject to introversion and tended to be more people oriented, tended, as a group, to be happier, were more self satisfied with regard to their professional and personal lives, and were more self confident. They also tended to have less college teaching experience, less solo competition at the high school level, and were somewhat younger and more satisfied with music education than the unsuccessful group (Lutz 1963:112-114).

He also found little agreement among the ratings. The successful group was rated highest by their fellow teachers and lowest by their students while the unsuccessful group was rated highest by their students and lowest by their administrators (Lutz 1963:133-134). He did not investigate these criterion ratings nor weight them for his composite rating criterion.

Lutz's study is important to this investigation. First, he found significant relationships between personality variables and instrumental music teaching success. Second, profiles between the two groups were quite similar, differences only in degree; yet, when compared to the general population, he found similar results to indicate that music educators were homogeneous as a group, quite different from the general population. Third, the criterion variables indicated a need to be investigated as well. This study was one of the few to use peer and pupil ratings.

His study is thorough and of value to the music education profession. While this investigator would have preferred to see some reliability and validity estimates for the investigator developed Opinion Ratings Form and the Instrumental Music Teacher Questionnaire, these instruments appeared useful for the study.

Fosse (1965) investigated biographical, demographical, experiential, environmental, and psychological factors to instrumental music teaching success, specifically high school band directors' Illinois State Music Festival rating. He stratified his sample into 4 groups, three groups of success and one group which did not enter the music contest (N=30). He investigated a total of 350 individual experimental variables from which 33 ordinal variables were found to be significant at the .05 level for his final discriminant analyses. He, too, found personality factors as measured by the MMPI to be the most significant (See Section One of this chapter).

Fosse's study is voluminous, a 536 page research project which found some important relationships. Like Lutz, whose study was similar, Fosse found MMPI variables important to discriminating among success groups of music educators, specifically high school band directors. His use of the discriminant analysis technique was admirable. Only one study since has used this technique to differentiate music education groups on various experimental variables (Whellams 1970). Lutz would have been better advised to use this procedure had it been available to him rather than taking each individual experimental variable and comparing significance between groups by chi square. However, with a total sample of only 30 subjects and 33 significant variables for use in the discriminant functions, Fosse erred on methodological grounds. Fosse emphasizes that the discriminant analysis program at Northwestern was new, and the developer wanted to test the program with a small sample of subjects. No where in this thesis is there any discussion of the resulting test of the effectiveness of the multidiscriminant analysis with a small sample in comparison to the efficacy of it with a large sample. One wonders whether this was a valid statistical treatment technique to use or whether Fosse would have been better advised to use analysis of variance or multiple regression techniques. He did not cross validate his findings probably due to the small sample size. Discriminant analysis requires a specific relationship of the independent variables to the number of subjects in the sample:

> Another rule is that the total sample size should be at least two and preferably three times the number of variables used . . . In fact, to be realistic, there are good reasons to insist that the size of the smallest group be no less than the number of variables used (Tatsuoka 1970:38).

Fosse should have used no more than 5 independent experimental variables because of only 5 persons in his smallest group. It is reasonable to assume that in 1965 when Fosse was completing his study the discriminant analysis program had not reached the level of sophistication to which it has today. Therefore, it may be imprudent to criticize him for selecting this fine technique which has not been used in music education research nearly enough. Nevertheless, Fosse's findings may be of dubious value. The significance of his study to this present investigation is that the discriminant analysis technique appeared useful for discriminating among the three success groups and nine success groups formulated for study. However, previous analyses would be required in order to assure the proper use of discriminant functions for predicting music teaching success.

The aforementioned research studies in music education as well as those discussed in Section One of this chapter have found personality variables to be significantly related to teaching success criteria. The decision was made, therefore, to concentrate on the personality variables alone as predictors of music teaching success.

BEST COPY AVAILABLE

Primarily, three prominent tests were used in the literature: the Minnesota Multiphasic Personality Inventory (MMPI), the Sixteen Personality Factor Questionnaire, (16PF), and the Guilford-Zimmerman Temperament Survey (GZTS). Others included the Edwards Personal Preference Survey (EPPS), the Ohio Psychological Examination, the California Psychological Inventory, and so forth. In music, however, the first three were the most used.

In selecting a test for this present investigation, effort was directed to locate tests which would survey the relative domains of personality and yet not require a detrimental amount of music teacher and music student teacher time. The MMPI includes 12 scales which relate to basic personality traits determined by six rigorous observational reports. The reliabilities and validities of the test are sufficient and its wide use certainly has indicated its popularity in the literature. However, the dimensions or behaviors measured refer to psychologically deviate areas. The test was empirically derived to separate normals from abnormals, persons in need of psychological treatment and counseling. Morgan, Duda, Lutz, Williams, and Fosse used one administration of the MMPI. Nevertheless, for the above reasons, it was decided not to use the pathologically oriented MMPI in this present study.

The GZTS measures 10 behavioral dimensions determined through factor analytic techniques. It is a good test. In music, however, only Anderson (1965) used it and three factors were significant. Both the GZTS and the MMPI have a Masculine-Feminine scale on which music educators register highly feminine. One must, therefore, question these tests on the grounds that they have not sufficiently defined the area of artistic or musical endeavors but rather classify persons on the feminine side if their values are artistic and on the masculine side if their values are mechanical and nonaesthetic. Therefore, that scale alone is of questionable value today, especially with the women's liberation movement. Furthermore, the scales tend to have positive and negative connections indicating right or wrong answers. But, one can question whether it is right or wrong to be high or low on certain scales.

The 16PF was used by Anderson (1968), Erickson (1954), Lamke (1951), and Funk (1973). Raymond B. Cattell, the author of the 16PF, has been one of the leading researchers in the area of personality and motivation measurements since 1920. The 16PF has been updated and has 16 basic primary factor traits that it measures along with 8 secondary and 4 tertiary or criterion derived scores. It appears to be the best test for measuring a broad spectrum of the personality trait domain.

Getzels and Jackson (1963) report only four research investigations that had used the 16PF test with groups of successful and unsuccessful teachers. They conclude their survey of the literature by stating:

> Too little has been done with the 16PF Test, with teachers as subjects, to accept without question the results that have appeared and to make possible a comprehensive evaluation of the method. Nevertheless, the instrument has at least two specific advantages (aside from purely technical considerations). First, by providing scores on factors that are not purely evaluative (i.e. psychologically "good" or "bad"), the test encourages the use of hypotheses that are more sophisticated than those linking "adjustment-maladjustment" or some such dichotomous variables to the complex phenomena of teaching and of teaching effectiveness. Second, the instrument derives from an extensive program of both theoretical and empirical work carried out by Cattell and his associates over a number of years . . . The resulting body of concepts and findings would seem of considerable heuristic value for investigators intending to use the 16PF Test for studies of teacher personality (Getzels and Jackson 1963:553-554).

One additional factor was that Cattell is not only a Professor of Psychology at the University of Illinois but that his organization, the Institute for Personality and Ability Testing (IPAT), is located in Champaign, Illinois. An extensive staff of skilled psychometricians is available for consultation and assistance. For the aforementioned reasons, the 16PF was selected as the data gathering instrument for the personality trait domain.

Studies in the literature that have included personality factors have used only one test administration of the same personality inventory to gather data. Personality tests have lower reliabilities than do achievement and aptitude tests. Therefore, if personality variables were to be the experimental variables under study, the best measurement procedure demanded more than one personality test administration to gather data. Since the investigator did not want to investigate differences among psychological tests but instead to sample a relatively large area of the psychometric domain, Cattell in communication to the investigator, suggested the use of both Forms A and B of the 16PF as well as the Motivation Analysis Test: Form A; the second form is still in the preparation stages.

The MAT Test measures the area of the dynamic domain - feelings, attitudes, beliefs, values, sentiments, and drives - which was considered to be important and complimentary to the personality trait domain; yet, no research study in music had attempted to measure this domain in depth nor measure the personality trait domain with any degree of robustness. The Minnesota Teacher Attitude Inventory (MTAI) was considered as were other interest tests by Kuder and Strong. However, these tests by their very nature could be faked as reported earlier in this chapter by Getzels and Jackson. Since MAT was new and was interesting for participants to take and provided 20 motivation variable scores, there was no question that this relatively new test was presently the best one available for measuring motivation. Furthermore, since interactions might be important, it was felt that this test might further define the behavior-trait domain. For example, intelligence has been a significant factor in practically every study. MAT aids intelligence differentiation by further delineating this broad trait. Three personality instruments were selected which comprised a maximum testing time of 3 hours. This provided the most comprehensive testing of music education subjects to date by any research investigation. The total sum of the related literature hinted strongly at the need for an in depth measurement of personality domains. To this writer, it appeared that previous studies merely tapped the reservoir finding significant droplets.

If one could enfilade the personality domains, a significant deluge might occur.

In summary, the literature in the area of personality being significantly related to music teaching and music student teaching success was affirmative. Depending upon the success criterion chosen by the investigator, predictions have been formulated to some degrees of usefulness. From this review, the following conclusions affected this present investigation.

1. Personality traits and behaviors can be adequately measured. Published tests are generally reliable and valid and an in depth measurement is desired.

2. The relationships, interrelationships, and interactions among personality, motivational, and criterion variables should be investigated first with inservice music educators for profile information useful in predicting success. The data from the music educator group could then be cross validated on the music student teacher group. This would provide some information useful to deciding whether success criteria and profiles are similar for both groups. Furthermore, the regression equations should also be cross validated on another group of music educators as well.

3. A multifarious sample for both experimental groups was desired. Since the music educator group would be located in diverse music teaching situations in the United States, the music student teacher group should include students majoring in music education at seven different teacher training institutions.

4. Multivariate statistical analyses would provide the most useful data in terms of relationships, interrelationships, and interactions. To make comparisons between groups using statistical treatments for each variable such as chi square or t-tests was deemed inappropriate and unsophisticated.

5. The music educator sample would need to be of sufficient size to allow for a certain among of attrition due to criterion data gathering. The goal was to obtain a committed sample of 300 music educators for personality testing.

BEST COPY AVAILABLE

REFERENCES

Albert, H. R., "An Analysis of Teacher Rating by Pupils in San Antonia, Texas," Educational Administration and Supervision, XVI, January 1941, 267-274.

Anderson, Harold M., "A Study of Certain Criteria of Teaching Effectiveness," Journal of Experimental Education, XXIII, September 1954, 41-77.

Anderson, James Harold, "The Predictive Value of Selected Factors for Student Teacher Success at Western Carolina University," unpublished doctoral dissertation, University of Tennessee, 1968.

Anderson, John Martin, "The Use of Musical Talent, Personality, and Vocational Interest Factors in Predicting Success for Student Music Teachers," unpublished doctoral dissertation, University of Southern California, 1965.

Barr, Arvil S., Theodore L. Torgerson, Carl E. Johnson, Vergil E. Lyon, and Anthony C. Walvoord, "The Validity of Certain Instruments Employed in the Measurement of Teaching Ability," The Measurement of Teaching Efficiency, ed. Helen M. Walker, New York: The MacMillan Company, 1935, 72-89.

Barr, Arvil S., Robert A. David, and Palmer O. Johnson, Educational Research and Appraisal, New York: J. B. Lippincott Company, 1953.

Barr, Arvil S., "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations," Journal of Experimental Education, 1-155.

Barth, George William, "Some Personality and Temperament Characteristics of Selected Music Teachers," unpublished doctoral dissertation, University of Southern California, 1961.

Benner, Charles H., "The Relationship of Pre-Service Measures to Ratings of Music Teaching Success," unpublished doctoral dissertation, The Ohio State University, 1963.

Bodanske, William Holt, "Selected Conditions Associated With The Mobility of Secondary School Music Educators in Missouri," unpublished doctoral dissertation, University of Missouri at Columbia, 1969.

Bodenstab, Ernest W., "Personality Patterns and Heterogeneity Among Teaching Interest Areas: A Multivariate Approach," AERA Paper Abstracts, 1968, New York Annual Meeting, 319.

Borkowski, Francis Thomas, "The Relationship of Quality of Work in Undergraduate Music Curricula to Effectiveness of Instrumental Music Teaching in the Public Schools," Journal of Experimental Education, Fall 1970, 39:14-19.

Brophy, Jere E., "Stability of Teacher Effectiveness," American Educational Research Journal, Summer 1973, 10:3, 245-252.

Bruner, Jerome S., "Process of Education Revisited," Phi Delta Kappan, 53:18-21, September 1971.

Bullock, Jack A., "An Investigation of the Personal and Professional Attributes of Superior Teachers of Junior High School Instrumental Music in New York State," unpublished doctoral proposal, University of Miami, June, 1972.

Bunyard, Robert Louis, "Factors Affecting the Transfer and Dropout of High School Instrumental Music Teachers," unpublished doctoral dissertation, University of California at Los Angeles, 1963.

Cady, Henry L., "Tests and Measures in Higher Education: School Music Teachers," *Journal of Research in Music Education*, XV, 2, 1967, 139-150.

Calder, Robert W., "Factors Influencing Male Music Education Graduates of Certain Pennsylvania Institutions of Higher Education to Leave or Not Enter the Profession," unpublished doctoral dissertation, The Pennsylvania State University, 1962.

Cattell, Raymond B., K. Barton, and T. E. Dielman, "Prediction of School Achievement from Motivation, Personality, and Ability Measures," *Psychological Reports*, 1972, 30, 35-43.

Colwell, Richard J., *Interpretive Manual - Music Achievement Tests I and II*, Chicago: Follett Publishing Corporation, 1969.

Colwell, Richard J., "Music Education and Experimental Research," *Journal of Research in Music Education*, XV, 1967, 1, 73-84.

Colwell, Richard J., *The Evaluation of Music Teaching and Learning*, Englewood Cliffs: Prentice-Hall, Inc., 1970.

Crews, E. Katherine, "The Role of the Supervisor of Student Teaching in Music - A Critique," Council for Research in Music Education *Bulletin* 20, Spring 1970, 33-37.

Cronbach, Lee J. and Lita Furby, "How We Should Measure 'Change'--Or Should We?" *Psychological Bulletin*, 1970, 74, 68-80. See also Errata, Inid., 1970, 74, 218.

Duda, Walter Boleslav, "The Prediction of Three Major Dimensions of Teacher Behavior for Student Teachers in Music Education," unpublished doctoral dissertation, University of Illinois, 1961.

Dvoral, Wayne, "An Investigation of the Colwell Music Achievement Test As An Instrument for Diagnostic Evaluation for Undergraduate Music Education Students," unpublished doctoral dissertation in progress, University of Illinois, 1973.

Durflinger, Glenn W., "Personality Correlates of Success in Student Teaching," *Educational and Psychological Measurement*, 1963, XXIII, 2, 383-392.

Dyer, H. S., R. L. Linn, and M. J. Patton, "A Comparison of Four Methods of Obtaining Discrepancy Measures Based on Observed and Predicted School System Means on Achievement Tests," *American Educational Research Journal*, 1969, 6, 591-605.

Edwards, Roger H., "Some Considerations For The Interfacing of the Arts and Evaluation," CEMREL internal paper to be published, March 1973.

Erbes, Robert Leonard, "The Development of an Observational System for the Analysis of Verbal Interaction in the Rehearsal of Musical Organizations," unpublished doctoral dissertation, University of Illinois, 1972.

Erickson, Harley E., "A Factorial Study of Teaching Ability," *Journal of Experimental Education*, 1954, XXIII, 1-39.

Ernest, David J., "The Prediction of Academic Success of College Music Majors," *Journal of Research in Music Education*, 1970, XVIII, 3, 273-276.

Farwell, Gaylord H., "Factors Influencing Male Music Education Graduates of Certain Pennsylvania Institutions of Higher Education to Leave or Not Enter the Profession - A Critique," Council for Research in Music Education *Bulletin*, Spring 1964, 3, 48-51.

Flanagan, Carroll Edward, "The Predictive Value of the Minnesota Multiphasic Inventory," *Journal of Experimental Education*, June 1961, XXIX.

Flanders, Ned A., *Analyzing Teaching Behavior*, Reading: Addison-Wesley Publishing Company, 1970.

Folstrom, Roger James, "A Comparative Study of the Musical Achievement of Students in Three Illinois High Schools," unpublished doctoral dissertation, Northwestern University, 1967.

Fosse, John Benjamin, "The Prediction of Teaching Effectiveness: An Investigation of the Relationships Among High School Band Contest Ratings, Teaching Characteristics, and School Environment Factors," unpublished doctoral dissertation, Northwestern University, 1965.

Fosse, John B., "The Relationship of Pre-Service Measures to Ratings of Music Teaching Success - A Critique," Council for Research in Music Education *Bulletin*, Spring, 1966, 7, 69-77.

Funk, William Russell, "Comparative Personality Variables, Professional Practices, Opinions and Attitudes in Southern Arizona High School Administrators and Teachers of Performing Groups in Vocal and Instrumental Music," unpublished doctoral dissertation, The University of Arizona, 1973.

Getzels, J. W., and P. W. Jackson, "The Teacher's Personality and Characteristics," *Handbook of Research on Teaching*, ed. J. L. Gage, Chicago: Rand McNally and Company, 1963, 506-582.

Glass, Gene V. and Julian C. Stanley, *Statistical Methods in Education and Psychology*, Englewood Cliffs: Prentice-Hall, Inc., 1970.

Gordon, Edwin, "Third-Year Results of a Five-Year Longitudinal Study of the Musical Achievement of Culturally-Disadvantaged Students," *Experimental Research in the Psychology of Music, Volume 8*, Iowa City: University of Iowa Press, 1972, 45-64.

Gray, John J., "Subject Matter Testing Practices in State University Music Education Curricula," unpublished doctoral dissertation, University of Southern California, 1960.

Guilford, J. P., *Psychometric Methods*, New York: McGraw-Hill Book Company, Inc., 1954.

Hellfritzsch, A. G., "A Factor Analysis of Teaching Abilities," *Journal of Experimental Education*, December 1945, XIV.

Hill, John D., "A Study of the Musical Achievement of Culturally Deprived Children and Culturally Advantaged Children at the Elementary School Level," *Experimental Research in the Psychology of Music VI*, Iowa City: University of Iowa Press, 1970, 95-123.

Holland, I. I. and R. I. Beezlay, "Personality, Motivation and Education Needed in Professional Forestry," *Journal of Forestry*, July 1971, 418-423.

Jones, Ronald D., "The Prediction of Teaching Efficiency from Objective Measures," *Journal of Experimental Education*, 1946, XV, 85-99.

Killackey, Jim, "Accountability Roots Planted," *The Sunday Oklahoman*, September 9, 1973, 23.

Krueger, Reynold J., "Extra-Curricular Activities and Music Education Students: A Pilot Study," unpublished research report, Department of Music Education, University of Illinois, January 1971.

Krueger, Reynold J., "The Oklahoma City University School of Music Testing Program for Freshman Students, 1972-73," unpublished research report, September 1973.

Lamke, Tom A., "Personality and Teaching Success," *Journal of Experimental Education*, December 1951, XX, 217-260.

Lee, Robert E., "An Investigation of the Use of the Musical Aptitude Profile with College and University Freshman Music Students," *Journal of Research in Music Education*, 1967, XV, 4, 278-288.

Lehman, Paul R., "The Predictive Measurement of Musical Success," *Journal of Research in Music Education*, 1969, XVII, 1, 16-31.

Lehman, Paul R., "A Selected Bibliography of Works on Music Testing," *Journal of Research in Music Education*, 1969, XVII, 4, 427-442.

Lehman, Paul R. and others, "School Personnel Research and Evaluation Services; Common Examinations; Music Education Examination," a critique in the *Journal of Research in Music Education*, 1972, 20, 3, 413-414.

Leonhard, Charles and Robert W. House, *Foundations and Principles of Music Education-Second Edition*, New York: McGraw Hill Book Company, 1972.

Lins, Leo Joseph, "The Prediction of Teaching Efficiency," *Journal of Experimental Education*, 1946, XV, 2-60.

Lord, Frederic M., "Elementary Models for Measuring Change," in *Problems in Measuring Change*, Chester W. Harris, ed., Madison: University of Wisconsin Press, 1963.

Lord, Frederic M., "A Paradox in the Interpretation of Group Comparison," Psychological Bulletin, 1967, 68, 304-305.

Lord, Frederic M., "Statistical Adjustments When Comparing Pre-existing Groups," Psychological Bulletin, 1969, 72, 336-337.

Lutz, Warren William, "The Personality Characteristics and Experiential Backgrounds of Successful High School Instrumental Music Teachers," unpublished doctoral dissertation, University of Illinois, 1963.

Madsen, Clifford K. and Charles H. Madsen, Jr., Experimental Research in Music, Englewood Cliffs: Prentice-Hall, Inc., 1970.

Maier, John Arthur, "An Investigation of the Relationships of Selected Academic and Musical Variables to the Two Year Basic Music Theory Criteria at the State University - Fredonia, New York," unpublished doctoral dissertation, The Ohio State University, 1970.

Marks, Edmond, and Charles G. Martin, "Further Comments Relating to the Measurement of Change," American Educational Research Journal, 1973, 10, 3, 179-192.

Michael, William B., George Barch, and Henry F. Kaiser, "Dimensions of Temperament in Three Groups of Music Teachers," Psychological Reports, 1961, 9, 701-704.

Morgan, Hazel Nohavec, "The Ability Pattern of Senior College Students Majoring in Music Education at the University of Minnesota," unpublished doctoral dissertation, University of Minnesota, 1946.

Nolin, Wallice H., "Patterns of Teacher-Student Interaction in Selected Junior High School General Music Classes," unpublished doctoral dissertation, The Ohio State University, 1969.

O'Connor, Edward F. Jr., "Extending Classical Test Theory To The Measurement of Change," Review of Educational Research, 1972, 42, 1, 73-97.

O'Connor, Edward F. Jr., Personal Communication, 1972.

McNeil, John D., and W. James Popham, "The Assessment of Teacher Competence," in Second Handbook of Research on Teaching, Robert M. W. Travers, ed., Chicago: Rand McNally, Inc., 1973.

Poland, William, "Measuring Musical Behavior and Predicting Success in the University School of Music," paper prepared for AERA Annual Meeting, February 14, 1963, Chicago, Illinois.

Popham, W. J., "Performance Tests of Teaching Proficiency: Rationale, Development, Validation," American Educational Research Journal, 1971, 8, 105-117.

Prince, Warren, "Fulfilling The Need for Replication in Music Education Research," Council for Research in Music Education Bulletin, 20, Spring, 1970, 1-6.

BEST COPY AVAILABLE

Rainbow, Edward, "Patterns of Teacher-Student Interaction in Selected Junior High School General Music Classes - A Critique," The Council for Research in Music Education *Bulletin*, 26, Fall, 1971, 44-46.

Rawls, Donna J. and James R. Rawls, "Personality Characteristics and Personal History Data of Successful and Less Successful Executives," *Psychological Reports*, 1968, 23.

Reimer, Bennett, *A Philosophy of Music Education*, Englewood Cliffs: Prentice-Hall, Inc., 1970.

Rhinehart, Charles B. Jr., "Incidence and Causes of Withdrawal from Public School Music Teaching," unpublished doctoral dissertation, Florida State University, 1963.

Schwartz, Anthony N., "A Study of the Discriminatory Efficiency of Certain Tests of the Primary Source Personality Traits of Teachers," *Journal of Experimental Education*, XIX, September 1950, 63-93.

Shanks, James Louis, "Concept Achievement in Science and Its Relationship to Some Non-Intellectual Characteristics of Prospective Elementary Teachers," unpublished doctoral dissertation, University of California at Berkeley, 1969.

Shatin, Leo, Wallace Kotter, and Gladys Longmore, "Personality Traits of Music Therapists," *Psychological Reports*, 1968, 23, 573-574.

Sidnell, Robert G., "A Comparative Study of the Musical Achievement of Students in Three Illinois High Schools - A Critique," Council for Research in Music Education *Bulletin*, 1971, 16, 60-65.

Silberman, Charles E., *Crisis in the Classroom: The Remaking of American Education*, New York: Random House, 1970.

Singer, Arthus, "Social Competence and Success in Teaching," *Journal of Experimental Education*, XXIII, December 1954, 91-131.

Sperry, Gale L., "A Study of Music Education Graduates from the University of Minnesota (1947-1955) with Particular Reference to Their Participation and Persistence in the Profession," unpublished doctoral dissertation, University of Minnesota, 1958.

Stake, Robert E., "To Evaluate an Arts Program," unpublished draft copy, University of Illinois at Urbana-Champaign, April 3, 1973.

Strub, Paul, "The Undergraduate Characteristics of Successful Public School Music Teachers," unpublished doctoral dissertation, University of Kansas, 1957.

Smith, Martha Lucile, "A Study of Elementary Student Teacher Confidence in and Attitudes Toward Music and Changes that Occur in a Student Teaching Experience," unpublished doctoral dissertation, Michigan State University, 1969.

Tatsuoka, Maurice M., *Discriminant Analysis: The Study of Group Differences* Champaign: Institute for Personality and Ability Testing, 1970.

Tatsuoka, Maurice M., *Multivariate Analysis: Techniques for Educational and Psychological Research*, New York: John Wiley and Sons, Inc., 1971.

Taylor, Corwin H., "Opinions of Music Teachers Regarding Professional Preparatory in Music Education," *Journal of Research in Music Education*, XVIII, 1970, 4, 330-339.

Turrentine, Edgar M., "Predicting Success in Practice Teaching in Music," unpublished doctoral dissertation, State University of Iowa, 1962.

Watkins, Robert H., "The Construction and Testing of a Forced Choice Scale for the Evaluation of Student Teachers in Vocal Music at the Junior High School Level," unpublished doctoral dissertation, University of Illinois, 1958.

Whellams, Frederick S., "The Relative Efficiency of Aural-Musical and Non-Musical Tests as Predictors of Achievement in Instrumental Music," Council for Research in Music Education *Bulletin*, 21, Summer 1970, 15-21.

White, Howard G., "Incidence and Causes of Withdrawal from Public School Music Teaching - A Critique," Council for Research in Music Education *Bulletin*, Fall 1965, 6, 44-46.

White, Howard G., "The Professional Role and Status of Music Educators in the United States," *Journal of Research in Music Education*, 1967, XV, 1, 3-10.

Williams, Raymond Edmund, "The Measurement and Predicting of Cooperating Teacher Effectiveness in Music Education," unpublished doctoral dissertation, University of Illinois, 1958.

Wilson, George Hugh, "A Study of Professional Music Education at Ohio State University," unpublished doctoral dissertation, The Ohio State University, 1958.

Wink, Richard L., "The Relationship of Self-Concept and Selected Personality Variables to Achievement in Music Student Teaching," *Journal of Research in Music Education*, XVIII, 1970, 3, 234-241.

CHAPTER III

PROCEDURES

In this research, an attempt has been made to investigate the relationships and interrelationships of personality and motivational variables to music teaching success criteria. The selection of the criteria in a study of this type may well affect the results of the study as may the methodological and statistical procedures. Some of the implied assumptions are: ". . . (1) There is empirical evidence of the relationship, and (2) the subsequent course does not alter in method, content, or students' learning characteristics (Bloom and others 1971:70). This study may define certain relationships either empirically or statistically, but if the nature of the teaching position changes (what Bloom and others refer to as alterations in method or content of subsequent courses), the relationships may be meaningless. While the teaching profession is changing, any prediction multiple regression results must be viewed cautiously. They add:

> It is our view that great care must be taken in using summative examination results for predictions regarding subsequent courses [future successes] and that technical help should be sought for the task . . . It is highly questionable whether a single test could effectively serve the purpose of prediction in combination with other summative evaluation intents (Bloom and others 1971:70).

In the present chapter, the sequential procedures and methodologies are recounted. Included are (1) data gathering, (2) data scoring, and (3) data anlyses.

DATA GATHERING

Data were gathered in two main areas: (1) experimental variables which included 16 personality temperament trait behaviors and 20 motivational dynamic behaviors, and (2) criterion variables which were used to define music teaching success in terms of pupil growth and ratings of teaching effectiveness. Background data about the participating music educators, music student teachers, their pupils, and their schools were also gathered.

Experimental Variables

The Sixteen Personality Factor Questionnaire, Forms A and B and the Motivation Analysis Test by Raymond B. Cattell and others were used to gather personality and motivational information. The tests and machine scored answer sheets were purchased from the Institute For Personality and Ability Testing. A total of 36 psychometric variables were investigated.

All tests were completed by the music educators during the period April, 1971 through March, 1972. Over 50 percent of the sample was group tested by the investigator. Others were mailed testing packets for completion at home and included a stamped, self-addressed envelope for return to the project office.

Because of ethical considerations and standards regarding personality and ability testing, the writer and the project consultant completed the forms as required by the American Psychological Association. Personality information is, by nature, an invasion of privacy. Therefore, the utmost caution was exercised in handling the data.

> Insofar as some tests of emotional motivational, or attitudinal traits are necessarily disguised the subject may reveal characteristics in the course of such a test without realizing that he is so doing. Although there are few available tests whose approach is subtle enough to fall into this category, the possibility of developing such indirect testing procedures imposes a grave responsibility on the psychologist who uses them . . . Of primary importance in this connection is the obligation to

> have a clear understanding with the examinee regarding the use that will be made of his test results (Anastasi 1968:33).

The personality tests used in this study were subtle and especially so in measuring the motivational dynamic domain. For this reason, meetings were held with the participants and the nature of the project was discussed. All data feedback were returned to the music educators' home addresses in order to assure confidentiality.

At the time of the initial meetings as well as at the group personality testing sessions, a background information sheet was completed by the music teachers and music student teachers. This facilitated the mailing of data feedback. The form gathered some necessary background information as well as participants' school and home addresses and telephone numbers. While most of the criterion tests were mailed to the schools for data gathering, results of the teachers' personality and criterion evaluations were mailed to their homes. Only the writer and his secretary had knowledge of individual music educators' and music student teachers' personality and motivation scores. Codes were used to assure confidentiality.

Criterion Testing

One of the major purposes of music education is to affect learning. If learning occurs, this achievement can be objectively measured and to some extent evaluated. Permission was obtained from the Follett Educational Corporation to use the Music Achievement Tests I-IV by Richard J. Colwell in this study. Because test validity in a pretest-posttest design was of real concern as far as measuring each music educator's pupils and using the class mean residual gain scores as constructs of music teacher accountability and success, it was necessary to obtain permission to duplicate test answer sheets and to dub tape recordings of the subtests. Because a multifarious sample of participating music educators was selected for this study, no single published music achievement test could be assumed valid for objectively evaluating the pupils' growth of each participating music educator.

During the initial meetings or contacts by letter, the participating music educators completed the "Music Objective Selection Sheet" which listed 15 music behavioral objectives which could be operationally measured by the MAT tests. Each objective related to a specific test section or subtest; however, Tests A and B and Tests F and G were later combined into one subtest each (AB and FG) because the investigator decided that the test items for these subtests were too few in number which could lower the reliabilities of the subtests. This reduced the selection to 13 music behavioral objectives from the 4 MAT tests.

Each participating music educator selected 5 music behavioral objectives from which the writer selected 3 for pre and posttesting. The published answer sheets were cut and spliced in order to prepare the master sheet of a special test for each music educator which evaluated his 3 objectives. This provided a criterion referenced approach to the music achievement testing because the music educator selected the objectives which he considered important and for which he taught. This special music achievement test master sheet was used to lithographically duplicate the test answer sheets for the pupils of each music educator. From the original 290 music educators selecting among 13 music behavioral objectives, there were 47 separate, specially prepared music achievement tests each including 3 test sections or subtests. A master tape recording including the 13 music subtests was dubbed from the published records; from this master tape, the separate 47 music achievement tests tapes were duplicated. A professional recording studio in Champaign, Illinois, dubbed these test tapes in order to assure a high quality sound recording.

No more than 30 minutes was required for music achievement testing sessions. This was because the participating elementary music teachers had class periods averaging 40 minutes.

Packages including test tape, answer sheets, the "Music Achievement Tests" sheet of test giving instruction, the "Additional MAT Comments" sheet, and the "Music Achievement Test (MAT) Information Sheet" which provided information about the music class being tested were mailed to all participating music educators during the period August to October, 1971. The music educators were instructed to give the pretest at their earliest convenience and to return all materials including the test tape, answer sheets, and data sheets to the project office in the stamped, self-addressed envelope.

After 16 weeks of instruction the packages were once again mailed to the participating music educators for posttesting. The 16 week interval was selected in order to approximate one semester's worth of instruction. Interval, however, became an extenuating variable because many music educators did not posttest at the desired time due to concert commitments, forgetfulness, and heavy schedules. When this serious design error became apparent, interval had to be treated as a variable affecting the gain scores even though numerous reminders were sent the participating music educators to encourage them to follow, as closely as possible, the 16 week interval requirement.

Among the 243 music educators who music achievement tested their pupils, intervals for instruction ranged from 6 to 33 weeks. As will be mentioned later in this chapter, this design problem was treated statistically.

Not all participating music educators pretested during the first few weeks of the school year. Therefore, another design error crept into this study necessitating a stratification of 4 subgroups of music educators due to pretesting time differences. Table 3 shows the 4 pretesting times and the instructional intervals. While interval could be treated statistically, no treatment of the pretesting time differences could account for the possible amount of prior learning differences that would accrue between a group that was pretested in September and another that was pretested in January. The

TABLE 3.--MAT PRETESTING TIMES

Interval	Sept.-Oct.	Nov.-Dec.	Jan.-Feb.	March	Number of music ed.
6	1	0	0	0	1
9	0	0	0	2	2
10	0	1	1	1	3
11	0	0	1	1	2
12	0	0	0	1	1
13	0	0	2	1	3
14	0	0	6	0	6
15	1	2	12	0	15
16	6	2	10	0	18
17	7	6	6	0	19
18	39	8	18	0	60
19	9	8	11	0	28
20	12	5	6	0	23
21	7	8	4	0	19
22	2	7	1	0	10
23	4	8	1	0	13
24	0	1	0	0	1
25	5	4	0	0	9
26	2	1	0	0	3
27	1	1	0	0	2
28	0	1	0	0	1
29	2	0	0	0	2
30	2	0	0	0	2
31	2	1	0	0	3
33	2	0	0	0	2
	94	64	78	6	243

January group had had a semester's amount of music instruction. Four music educators used second semester classes that met for only 9 weeks of instruction. Therefore, pre and posttesting occurred throughout the 1971-72 school year.

One school district was already using the Music Achievement Tests at the time of this study in a pretest-posttest design. The music administrators graciously consented to follow an 18 week interval rather than a 36 week interval as they had planned. Since they had their own test answer sheets and scoring services, and gave the test district wide over television, their research division provided the investigator with the results of the participating music educators' pre and posttest pupils' scores. Because this was a city wide test administration in elementary classroom music, test validity was assured because the music education staff had formulated this testing procedure. It was not superimposed on them by the music administrators. This group used 4 subtests from the MAT tests which were incorporated in this study. In addition, three participating music educators elected to use one single test (MAT III) which provided 4 subtest scores. As in the case of the large district, a single total score was used for deriving the class mean residual gain scores. The only difference was that most of the music educators were evaluated on 3 objectives and 16 were evaluated on 4 objectives. In addition, one music educator failed to give the third subtest and thus only 2 subtests were used to derive his class mean residual gain score. This procedure was deemed appropriate because these music educators selected the subtests from the MAT series most appropriate for evaluating their music teaching goals.

Music educators teaching at the high school level were instructed, if possible, to use a less skilled group of pupils for music achievement testing than their best performing group. For example, a high school choral director was encouraged to use a tenth grade mixed chorus, theory class, or some other choral organization rather than his best, most elite group. Research as indicated in Chapter II of this investigation, tended to support the thesis that while performing groups learn and this achievement can be measured, regression effects could seriously affect the music educator's ability to produce measurable music learning with already sophisticated pupils.

Ratings of Music Teaching Success

The rating instrument selected for use in this investigation was the Illinois Teacher Evaluation Questionnaire (ITEQ). Ratings were gathered from the music educators' pupils, building principals, music supervisors, peer teachers, and themselves. The music educator selected the class for pupil ratings and was encouraged to use his best class, the one which he considered to be the most advanced musically, and the one in which he had developed the best interpersonal relationships. The purpose was for the music educator to have the best opportunity for obtaining successful pupil rating evaluations.

In other words, the investigator purposely built into the research design a strong possibility for pupil rating bias because the rating evaluations were under investigation. If pupil evaluations were to be meaningful, they needed to be obtained from a music class in which most of the pupils had a strong, intrinsic desire to achieve in music. Furthermore, the music educator was encouraged to have someone other than himself administer the pupil rating evaluations. In this type of evaluation, pupils' opinions and attitudes about their music teacher were gathered in order to obtain a type of pupil process measurement.

> . . . if these types of affective questions are asked in a context removed from the course in question, the validity of the responses is increased. Thus, questionnaires designed to measure attitudes toward English [or music] are more trustworthy when administered by a neutral person (for instance, a guidance counselor or homeroom teacher) than when administered by the English teacher [or music teacher] who wishes to have this kind of summative curriculum information (Bloom and others 1971:239).

It was with the above in mind that the music educators were encouraged not to administer the pupil rating evaluations themselves. A total of 206 music classes completed the pupil ratings. These were administered as follows:

7	building principal
49	another teacher
12	counselor
14	student
13	another person
96	participating music educator
16	unknown administrator
206	

A total of 110 music classes had someone other than the music educator administer the pupil rating evaluations. This represented 53 percent of the sample.

Packages of ratings were mailed to the participating music educators in May, 1972. Each package included ITEQ answer sheets for one class of pupils, a letter, the "Illinois Teacher Evaluation Questionnaire (ITEQ)" instruction sheet for giving and administering the ratings, and the "Music Class Description Form (ITEQ)" which gathered information about the music class used for the

BEST COPY AVAILABLE

rating evaluations. Also included in the package were 5 peer ratings each in a stamped, self-addressed envelope for the music educator to give to 5 of his peer teachers in his building who knew him well and a "Peer Teacher List" on which the music educator listed the names of the 5 peer teachers to whom he gave the peer rating envelopes. The "Peer Teacher List" facilitated follow up procedures, if needed; and this blank was also used to discourage any music educator from completing the peer evaluations himself. These peer teachers represented both music and other subject matter specialists. The package also included an envelope containing the "Self Rating" and a return package which was stamped and addressed for mailing to the project office.

Follow up letters along with telephone calls assisted in gathering these data. Only one music educator obtained the pupil evaluations in September, 1972.

Building principal and music supervisor rating forms which were stamped "confidential" were mailed directly to these persons and included a stamped, self addressed envelope for return to the project office. Furthermore, a letter explaining the project was once again included along with the "Illinois Teacher Evaluation Questionnaire Administrator Rating" blank which gathered some background information about the administrator and his school. He was requested to indicate whether his school was socially advantaged or socially disadvantaged with no middle choice category on the form. When the administrator indicated "a little bit of both" or some similar comment, a middle category became necessary. Furthermore, when the building principal indicated socially disadvantaged and the supervisor indicated socially advantaged, or vice verse, the school was placed in the middle category. While practically all data were gathered by the end of June, 1972, one additional follow up occurred in September, 1972, to building principals and music educators. While Chapter IV includes details on the numbers of music educators teaching in the three socioeconomic classifications, the total derivation sample (N=209) included

22 percent in socially disadvantaged schools, 15 percent in the middle combination category, and 63 percent in socially advantaged schools.

The music student teachers were rated by one class of pupils, their cooperating teachers, supervising teachers, peer music student teachers, and by themselves. These evaluations were gathered upon conclusion of the student teaching experience. Packages similar to those used for the music educator group were mailed to the cooperating teachers for gathering the pupil and cooperating teacher ratings. The teachers were instructed to use the class with whom the music student teacher worked best. A stamped, self addressed envelope was enclosed along with a class information sheet. Student teacher peer and self ratings were gathered in meetings with the music student teachers. The college supervisors were mailed the supervisor rating forms.

It was desired that these music student teacher evaluations would approximate those ratings gathered for the music educator sample. In only a few cases did the cooperating teachers refuse to assist this project because of the pupil rating factor which some interpreted to reflect on their abilities as cooperating music teachers because of the "Interest to the Class" subscore.

SCORING AND TABULATION OF DATA

Personality Tests

After gathering the music educators' and music student teachers' personality test data, the machine scored answer sheets were placed in a locked file cabinet. The music student teachers were filed by institution and name. For all 290 participating music educators, a separate file by school district, teacher's name, district's code number, and teacher's code number was maintained. As soon as all experimental data were in the project office and filed, the test answer sheets were transported to the Test Services Division of the Institute for Personality and Ability Testing for scoring. The scoring procedures provided a computer profile printout both for the 16PF and Motivation Analysis Test in duplicate - the original to be mailed as feedback to each participant and the copy to remain in the project office in case the original was lost in the mails. The scoring procedures also provided IBM data processing cards for the analyses. Because only the primary factors of the 16PF test and the 10 Integrated and 10 Unintegrated variables from the MAT test were under investigation, the 6 data processing cards provided by IPAT were reduced to 2 cards per teacher by the investigator using a computer data manipulation program.

The computer profiles as generated by IPAT for each participant were filed for later feedback in Fall, 1972. Copies of the 16PF and MAT data cards were duplicated so that one data set was stored at the project office, a second data set was stored at the investigator's home, and a third data set was stored at IPAT. This satisfied his paranoia about loss of data due to extraneous variables such as fire, theft, bombing, and so forth.

Pupil Music Achievement Tests

The pupil music achievement tests sheets were scored by hand by the investigator, his parents, and his secretary. Scoring accuracy was checked by the project consultant who randomly selected 41 test answer sheets from the 23,596 available. A scoring accuracy percentage was thus obtained. A "Music Achievement Test Data Sheet" similar to "Fortran Coding Form" sheets was lithographically duplicated so that each music pupil could be listed with the following data: last name and first initial, grade, pre and posttest music achievement scores for each subtest, and an 8 digit code number of which the first 4 digits identified the music educator's code number and the last 4 digits the pupil's number. Once these data sheets were complete for each music educator, data processing cards were keypunched and verified by the keypunch operators at the University of Illinois' Digital Computer Laboratory. This procedure generated 1 card per pupil for each participating music educator for a total of 11,798 cards.

In order to provide feedback to the participating music educators, a computer program was used which printed out the information on the cards. With an explanatory letter, pupil music achievement test scores and subtest norm sheets were thus provided as feedback to the music educators.

Class pre and posttest mean score cards for each teacher's total test were generated from the pupil cards containing their raw pre and posttest subtest scores. This produced 1 card per teacher of his class' mean pre and posttest music achievement scores. These music achievement data mean cards were then ready for computation of the class mean residual gain scores. This technique was selected to be the best one available for producing gain or difference scores and was discussed in Chapters I and II.

IBM data sheets were used to code the participants' background information as well as knowledge and information about the classes used for music achievement testing, interval between pre and posttesting, number of class sessions per week, number of minutes per class session, and rating information. One card per teacher was keypunched to facilitate computer analysis.

In order to validate the usefulness of the constructed G criterion, it was necessary to investigate differences in class mean residual gain scores that could be attributed to differences in pretesting times. Analysis of Variance (ANOVA) and t-test procedures were used to investigate differences among the following 4 pretesting groups: (1) September and October, (2) November and December, (3) January and February, and (4) March. Teachers were rank ordered according to the G criterion within each pretesting group that was statistically significant. These teachers were further stratified into rank ordered teacher type groups: female choral, female instrumental, male choral, and male instrumental. For differences that were found among the pretesting time groups, a percentage-ratio technique was used to rank order the subjects.

Subjects were selected from the rank ordered pretesting time groups and assigned a single ranked number according to the percentage of subjects within each group as far as the total group was concerned. For example, if all four groups were found to be statistically significantly different in G beyond the .05 level and this gain could be attributed to the differences in pretesting times, the following procedure was used to select a single rank ordered sample: persons from each of the 4 groups were assigned a rank according to the percentage of subjects in each pretesting teacher type group. If there were 3 times as many subjects in pretesting Group One than Group Four, the ratio was 4 to 1. This procedure provided a single rank ordered sample according to G as far as levels of goodness or success were concerned. If no differences among the pretesting groups were found beyond the .05 level of significance, the sample would be rank ordered according to G for all 209 subjects.

All of the statistical analyses conducted in the study adopted the .05 level of significance for rejecting or failing to reject the null hypotheses. Many of the actual figures are reported for clarity to the reader.

Because the investigator anticipated the possibility that at least one pretesting time group would be significantly different from the others and because correlational analyses would be undertaken among the gain and 5 rating variables, it was necessary to standardize the sample according to gain in a manner that was fair and equitable as found by the ANOVA and t-test analyses. Therefore, the music educator ranked gain scores were forced into a normal curve within each subsample of teacher type (instrumental and choral) and for each significant pretesting time group (1) September-October, (2) November-December, (3) January-February and (4) March. The derived class mean residual gain scores per se could not be used because of possible bias due to differences in pretesting times.

The subjects were assigned a number from 6 to 1 which represented the normal curve of 3 standard deviations above and below the mean ranked score for the teacher type, pretesting time group: the top 3 percent of the sample received a score of 5 while the bottom 3 percent of the sample received a score of 1. With this technique, the class mean residual gain score rank was standardized over the normal bell shaped curve. If no significant differences were found, the class mean residual gain scores would be standardized using a standard scores program to a mean of 5.5 and a standard deviation of 2.00. Either approach would facilitate the use of the criterion gain data in the correlational analyses.

Ratings of Music Teaching Success

The principal, supervisor, peer, and self ratings were coded with the teachers' code and school district numbers before mailing to the raters. Pupil ratings for both groups of participants (music educators and music student teachers) were scored at the Office of Measurement and Research, University of Illinois. Four subscores plus a total score were generated including a printout for teacher feedback and data processing cards. The printout compared the pupil evaluations of their music teachers to national teacher norms for all 40 test items, 4 subscores, and total score. These norms were based primarily on subject matter teachers and not music educators and may not be valid for a sample of music educators.

As the other ratings were received in the project office, these were filed in the participants' folders. After all the ratings were received, the information on the amended ITEQ, 23 item form was transferred to data processing cards by machine scoring techniques.

The investigator used a data manipulation program to generate a mean rating score for each category of rating: pupil, building principal, music supervisor, peer teacher, and self; and for the music student teachers: pupil, cooperating teacher, supervising teacher, peer music student teacher, and self. If more than one principal rated the music educator or more than one cooperating teacher rated the music student teacher, like the peer ratings, these were first averaged in order to derive one mean rating for each category. Each categorical rating card was kept separate providing 5 rating cards per subject.

The ITEQ rating form provides a 4 point scale, and the rater responds to statements in a forced choice manner: strongly agree, agree, disagree, and strongly disagree. Half of the items are worded positively while half are negatively worded statements.

The ratings in each of the 5 categories for both samples were standardized to a 10 point scale with a mean of 5.5 and a standard deviation of 2.00. This facilitated generating the R criterion.

Complete rating evaluations in all 5 categories were received from 183 music educators. There were 45 subjects who were not employed in systems with a music supervisor. Rather than use some sort of mean rating score to substitute for missing data, multiple regression equations predicted 1 or 2 missing ratings for those participating subjects with missing data.

Only 35 music student teachers had complete rating data, so multiple regression equations predicted up to 2 missing ratings for 38 subjects. Those with 3 or more missing ratings were eliminated from the study thus reducing the sample from 101 to 73 music student teachers.

All ratings that were available were included in the multiple regression equations. Even if a music educator could not be included in the final analyses of the study because of missing music achievement data while his rating data were complete, these data were used to predict missing data for other subjects. The total sample of 290 music educators was used for predicting missing data.

Once the 5 ratings per teacher were complete and missing data had been predicted, it was necessary to investigate differences in the 5 rating criterion variables (pupil, building principal, music supervisor, peer teacher, and self) that could be attributed to music teacher type (instrumental or choral) and grade level (upper elementary, middle, and secondary). Analysis of variance and t-test techniques were used. Teachers were stratified into the following 6 groups: (1) elementary choral, (2) elementary instrumental, (3) middle school choral, (4) middle school instrumental, (5) high school choral, and (6) high school instrumental. Since the individual ratings were already standardized across the sample, the teachers' 5 rating cards were used in these analyses. This procedure was deemed crucial because it was unknown whether upper elementary choral teachers, for example, would be rated differently in the 5 rating categories than other choral or instrumental music educators. Furthermore, ITEQ was designed for secondary school use; but many music educators in the study obtained pupil evaluations from upper elementary and middle school pupils. Differences among the ratings needed to be investigated first before producing the R criterion score for each participant.

The composite R criterion was generated through factor analytic techniques. The weights from the first principal component were used to obtain factor scores which, in turn, provided a single composite rating (R) criterion score per subject. These factor scores are, in effect, weighted averages, weighted according to the factor loadings.

> In any case, the method [second order factor analysis which generates factor scores], though not extensively used in the past, has great potential for complex behavioral research. Instead of using many separate test scores, fewer factor scores can be used . . . thus achieving considerable persimony and increasing the reliability and validity of the measures (Kerlinger 1973:677-678).

This weighted factor rating score (R) was used as a subjective criterion of music teaching and music student teaching success.

The above procedures in scoring, tabulating, and analyzing data provided four criterion variables for use in the study: G (gain), R (rating), G+R (gain plus rating), and G-R (gain minus rating). Teachers were rank ordered within the G and R criterion categories. The G+R category simply added the R ranked score to the G ranked score for a composite music educator criterion score while the G-R category subtracted the R ranked score from the G ranked score for a discrepancy criterion score.

The music student teachers were rank ordered according to R and used only as a cross validation sample. No separate analyses were conducted on the music student teachers as far as predicting their success from personality and motivational variables.

Twenty percent of the music educator sample served for cross validation purposes. This sample was selected randomly from the 4 teacher type groups so as not to be overweighted by any single group of music educators. The derivation sample of 168 subjects and the cross validation sample of 41 subjects were thus stratified.

DATA ANALYSES

Criterion Data

The first subproblem of the study asked: What are the relationships among teacher accountability criteria used to objectively define music teaching success?

a. class mean residual gain scores of pupils
b. teacher ratings by pupils
c. teacher ratings by principals
d. teacher ratings by supervisors
e. teacher ratings by peers
f. teacher self ratings

The previous discussion of procedures for generating the G and R criterion scores emphasized some preliminary analyses. Once the G, R, G+R, and G-R criterion variables were formulated, the following analyses were undertaken.

1. The ranked data for both G and R were investigated for significant relationships with the Spearman Rho technique, a rank correlation coefficient which is a variant of the Pearson Product Moment Correlation Coefficient. The hypothesis investigated was that there was no significant relationship between G and R.

2. Correlations among the standardized gain scores (6-1), sex, and the 5 ratings of music teaching success were undertaken to investigate the null hypotheses that no significant relationships existed.

3. Regular Pearson Product moment correlations were undertaken to investigate the null hypotheses that no significant relationships existed among the G, R, G+R, and G-R criteria. Correlational analyses were conducted for the total derivation sample (N=168) as well as for within teacher type groups: female choral, female instrumental, male choral, and male instrumental.

If there were significant relationships among the music teacher accountability criteria, a single composite criterion such as G+R could be used for prediction purposes. If there were no significant relationships, separate equations would be needed for predicting music teaching success. Therefore, the correlational analyses were crucial to the study as far as predicting music teaching success from personality and motivational variables were concerned.

While it is true that anlysis of variance and t-test statistical procedures might have been better for investigating hypotheses that there were no significant differences among the accountability criteria (these techniques were used before deriving the G and R criterion scores), the investigator decided to follow standard procedures.

Relationships and Interrelationships

The second subproblem asked the question: What are the relationships and interrelationships among the personality variables measured by the Sixteen Personality Factor Questionnaire and the motivational variables measured by the Motivation Analysis Test to music teaching success?

BEST COPY AVAILABLE

> Behavioral research is being revolutionized by multivariate thinking and analysis. Multivariate analysis is a general term used to describe a group of mathematical and statistical methods whose purpose is to analyse multiple measures of N individuals: multiple regression analysis, multivariate analysis of variance, canonical correlation, discriminant analysis, and factor analysis . . .
>
> The phenomena we wish to explain and predict are complex: achievement, learning, aggression, intelligence, creativity, risk-taking, organizational productivity, group cohesiveness, and many others. Many variables influence such phenomena, and multivariate methods are ways of studying multiple influences of several independent variables on one or more dependent variables . . .
>
> Multivariate methods . . . make it possible for the behavioral scientist to probe more deeply and more realistically into phenomena. The influence is profound: the very nature of the problems that behavioral scientists study is changing radically (Kerlinger 1973:602).

Personality and motivational dimensions of behavior are not necessarily discrete traits or variables which persons either possess or do not possess but rather are interrelated in such a way as to become a person's behavioral profile. Furthermore, these psychological dimensions may interact differently when analyzed in conjunction with criterion or success variables. A multivariate approach must be used because human nature is multivariate.

> In applied contexts, particularly in educational and psychological research, multivariate analysis is concerned with a group (or several groups) of individuals, each of whom possess values or scores on two or more variables such as tests or other measures. We are interested in studying the interrelations among these variables, in looking for possible group differences in terms of these variables, and in drawing inferences relevant to these variables concerning the populations from which the sample groups were chosen (Tatsuoka 1971:1).

The first analysis investigated the null hypothesis that there were no significant interrelationships among the experimental personality and motivational variables, the G and R criterion groups (the teachers were stratified into 9 success groups by dividing each of the G and R samples into thirds: high, medium, and low so that the 9 groups were high gain-high rating, high gain-medium rating, and so forth to low gain-low rating), and sex. A 9 by 2 by 36 multivariate analysis of variance (MANOVA) was employed to test this hypothesis.

The next analysis investigated the null hypothesis that there were no significant interrelationships among the experimental variables (predictive) and the 4 constructed criteria (G, R, G+R, and G-R) for each of the following 4 subsamples - male instrumental, male choral, female instrumental, and female choral - as well as between the male and female groups. Correlational analyses were used.

The correlational analyses were, in essence, looking for interactions between sex and personality variables and between teacher type and personality variables. If these correlations were somewhat opposite within the criterion domain, these interactions would show up as differences in the predictor-criterion relationships. For example, if males showed a positive correlation on Factor B (intelligence) to each of the 4 criteria while females showed a negative or less positive correlation on the same variable, an interaction could be suspected. The same would hold true for differences in the predictor-criterion relationships for teacher types. A two variable interaction such as sex and a criterion is a first order interaction while a three variable interaction such as sex, a personality variable, and a criterion variable is a second order interaction.

> Though infrequent compared to significant main effects, significant first-order interactions occur frequently enough to warrant serious attention. Students of research must be able to handle them. In the opinion of some behavioral researchers, especially in education, the study of interactions is becoming increasingly important and should become a central concern of researchers . . . The Cronbach and Snow report is definitive - and discouraging. They found the research of ability-trait interaction (ATI) to be frustrating and inconclusive. There are enough first-order interactions now appearing in both psychological and educational research, however, to be encouraging (Kerlinger 1973:257).

The above quote is especially pertinent to the study because of the nature of personality and motivational variables. Allowing for the fact that significant interactions might be found in the above analyses, these would need to be handled in at least two ways for predicting music teaching success: (1) predict success within each subgroup of music teacher type and sex which is the most ideal approach, or (2) generate additional cross product predictors (moderating variables), which would account for the interactions statistically. Both procedures were used. The investigator recognized that these were merely procedures for "getting around" interaction when it occurs.

> Certain variables that are known or suspected to influence achievement, such as sex or social class, may at the time be merely distracting complications. But, they must be controlled. We can "control" them by building them into the research design. Not only can they be controlled; they can yield information of possible value and significance (Kerlinger 1973:257).

The moderating variable representing the interaction (different relationships to the criterion for either teacher type and sex groups) was introduced into the regression equation as an additional predictor moderating variable, moderating because it was controlling statistically the interaction effects. This was created by multiplying the individual's score for the experimental variable times the dichotomized interaction variable, either sex or type, for each subject. Male subjects were coded 1 and females 2 while instrumental teachers were coded 1 and choral 2.

Multiple regression equations were formulated for the total derivation sample using the personality, motivational, and moderating variables as predictors to the 4 criteria (G, R, G+R, and G-R). These equations were cross validated on the cross validation sample of 41 subjects. "It can be said, I think, that multiple regression analysis and factor analysis are two of the most powerful, general, and useful modes of analysis available to the behavioral scientist" (Kerlinger 1973:602).

The general idea of cross validation is to check out the multiple regression equations derived from one sample of subjects on another sample of similar subjects. This procedure further validates the usefulness of the multiple regression data.

The 16PF and MAT (Motivation Analysis Test) variables were kept separate for the regression analyses due to the nature of their psychometric domains. The 16PF test measures the temperament-trait domain of intelligence, surgency, shyness, and so forth while the MAT test measures the dynamic domain of interests, values, drives, and sentiments. Therefore, these tests lie in two theoretically different psychometric domains.

For cross validation, the b-weights derived from the regression equations on the derivation sample were applied to the scores of the subjects in the cross validation sample. Correlations between the derivation and cross validation samples indicated the degree of usefulness and significance of the equations and cross validation. R^2 indicated the percent amount of variance of the criterion which could be determined or predicted by the linear combination of the predictor experimental variables. This is the coefficient of determination. R' is an approximate estimate of how much the multiple R would be lowered had the weights been applied to another like subject. Cross validation is a better procedure and was used in the study.

The above analyses investigated the relationships and interrelationships of personality and motivational variables to music teaching success criteria, the major purpose of the study, and were classified as Phase I analyses. While there were no indications in the literature to reinforce the investigator's suspicions that higher multiple correlation coefficients might be possible by stratifying the sample into music teacher type-sex groups, Phase II was undertaken in order to further identify these relationships and interrelationships.

Phase II

Multiple regression equations were formulated for predicting music teaching success (G, R, G+R, and G-R) from the personality and motivational variables for the following 4 subgroups of music educators: female choral, female instrumental, male choral, and male instrumental. These equations were cross validated to the music student teacher samples which were similarly stratified into music major groups on the R criterion. Because the sample of 209 music educator subjects was reduced into these aforementioned subgroups, no music educator teacher type cross validation group procedures could be employed. The sample simply became too few in numbers of subjects within subgroups. The equations were cross validated to each group, however, by applying the weights derived on one subgroup to the other 3 subgroups. This procedure investigated sample heterogeneity.

Personality Profiles

The third subproblem asked the question: Do personality profiles differ for successful and unsuccessful male and female music educators? This question is much more descriptive in nature than the second subproblem which, in essence, predicted a criterion from personality and motivational variables. In prediction, the approach is similar to following a recipe: a certain amount of experimental variable 1 is added to a different amount of experimental variable 2 and so on through all the experimental variables that contribute to the equation in such a way as to significantly predict a success criterion. Profile data, per se, cannot be inferred from the variables contributing to the regression equations.

The music educator sample (N=209) was stratified into successful and unsuccessful male and female groups for both the G and R criteria. The upper and lower 25 percent of the subjects were used: the middle 50 percent were not included in the profile analyses.

Two group stepwise discriminant analyses were conducted for each criterion and investigated profile differences for successful and unsuccessful male and female music educators: (1) successful males and successful females, (2) unsuccessful males and unsuccessful females, (3) successful males and unsuccessful males, and (4) successful females and unsuccessful females. The raw scores for the 16PF test were used because the sten scores account for sex differences while the sten scores for the Motivation Analysis Test do not account for population sex differences. A total of 16 profile analyses were investigated.

Profiles have been used in previous studies and have been, in fact, the most common way of investigating personality trait or motivational characteristics between groups. The usual approach has been to calculate the means and standard deviations for each group and conduct variable by variable analyses between the groups using t-test or chi square techniques. But, these approaches do not consider the interrelationships among the experimental variables. Tatsuoka notes the discrepancies between conclusions drawn from discriminant profile analyses and those based on mean differences on factor taken one at a time and concludes:

> Thus, examining the pattern of weights gives us a much more accurate account of the nature of group differences in terms of a given set of variables than does looking at each variable separately with no regard for their interrelations and partly overlapping information . . . This is because the weights are analoguous to the partial regression weights in a multiple regression equation, so each weight represents the relative importance of that variable with the effects of the other variables "partialled out" (Tatsuoka 1969:4).

The approach of constructing a linear combination reduces the problem of group differences to a univariate problem. Because two group profile comparisons were desired, the stepwise multiple regression technique was used which is equivalent to a two group discriminant analysis with one discriminant function. The b weights are analoguous. It was desired that profile data in conjunction with the regression data would assist in advising and counseling undergraduate music education majors, the fourth subproblem.

REFERENCES

Anastasi, Anne, Psychological Testing - Third Edition, Toronto: The MacMillan Company, 1968.

Bloom, Benjamin S., J. Thomas Hastings, and George F. Madaus, Handbook On Formative and Summative Evaluation of Student Learning, New York: McGraw-Hill Book Company, 1971.

Campbell, Donald T. and Julian C. Stanley, "Experimental and Quasi-Experimental Designs For Research on Teaching," in Handbook of Research on Teaching, N. L. Gage, ed., Chicago: Rand McNally and Company, 1963, 171-246.

Kerlinger, Fred N., Foundations of Behavioral Research - Second Edition, New York: Holt, Rinehart, and Winston, Inc., 1973.

Sidnell, Robert, "The Dimensions of Research in Music Education," Council for Research in Music Education Bulletin, 1972, 29, 17-27.

Tatsuoka, Maurice M., Multivariate Analysis: Techniques for Educational and Psychological Research, New York: John Wiley and Sons, Inc., 1971.

Tatsuoka, Maurice M., Validation Studies: The Use of Multiple Regression Equations, Champaign: The Institute for Personality and Ability Testing, 1969.

CHAPTER IV

ANALYSIS OF DATA AND FINDINGS

The findings of the study are reported in the chronological order in which the investigation was conducted. Discussed first is sample attrition. The criterion data are discussed second. The findings of the relationships and interrelationships among the criterion data leading to the formulation of the 4 criterion variables are presented. The third section includes the findings of the multiple regression equations used to investigate the relationships and interrelationships of personality and motivational variables to music teaching success criteria in terms of G (gain), R (rating), G+R (gain plus rating), and G-R (gain minus rating). The fourth section highlights the personality and motivational profiles for successful and unsuccessful groups of male and female music educators while the final section discusses the use of the findings for advising and counseling undergraduate music education students.

SAMPLE ATTRITION

Teachers' data were not used in the final analyses if they failed to satisfy one or more of the following design requirements: (1) pre and post-testing of their pupils, (2) more than 12 pupils for music achievement testing, and (3) more than 2 rating evaluations. Subjects were also eliminated from the study if their data were suspect in any way.

Campbell and Stanley cite the possibility of volunteer bias which they call the selection-specificity effect which is relevant to the internal validity of a research investigation. They emphasize the fact that volunteer subjects may be nonrepresentative of the population as a whole because they may be more secure in their teaching positions, have less fear of being inspected, and will not be afraid to be compared to other educators. They suggest:

> To help us judge on these matters, it would seem well for research reports to include statements as to how many and what kind of schools and classes were asked to cooperate but refused, so that the reader can estimate the severity of possible selective biases. Generally speaking, the greater the amount of cooperation involved, the greater the amount of disruption of routine, and the higher our refusal rate, the more opportunity there is for a selection-specificity effect . . .
>
> Often stringent sampling biases occur because of the inertia of experimenters who do not allow a more representative selection of schools the opportunity to refuse to participate . . One way to increase it [sampling representativeness] is to reduce the number of students or classrooms participating from a given school or grade and to increase the number of schools and grades in which the experiment is carried on (Campbell and Stanley 1963:189).

In following Campbell and Stanley's advice, the investigator decided to include as many music educators as possible in the study even though there might be only one participant per school district. This was the first phase of sample selection and the reason for meeting with persons who were attending inservice classes, summer school sessions, or workshops. While the majority of these persons were teaching in Illinois, a substantial number of subjects were teaching throughout the United States accounting for the final sample representing 80 school districts in 16 states. The investigator estimated that due to the nature of the study and in order to obtain a final usable sample of 200 participating music educator subjects, more than double that number would need to be invited to participate. For this reason, over 500 music educators were invited to participate.

The attrition from over 500 invited to 290 committed subjects accounting for 58 percent of the sample population was due primarily to two factors: (1) the intensive testing of each subject with three psychometric instruments which may have been viewed as an invasion of privacy, and (2) the extensive amount of gathering criterion data in terms of pupil growth and ratings of music teaching success. Time was also an important factor contributing to potential subjects' refusing to participate in the study.

Refusal information was conveyed to the investigator in person, by telephone conversations with directors of music education and music teachers, and by letter. One very large school district on the west coast did not permit its music educators to participate because the director of research told the music administrator (who wanted to assist the study) that he had "been stung in the past with personality studies."

Other pressures offered as rejection reasons included newly instituted busing schedules for achieving integration of the schools, new evaluation methods and procedures being tested in the schools, and a general bureaucratic problem of obtaining administrative approval. Many administrators did not want to put forth the effort when no immediate gain to them would be apparent.

A second type of attrition occurred in the study which reduced the sample from 290 to 209 music educators with sufficient criterion data (28 percent attrition). The data about secondary attrition came to light during the criterion data gathering, scoring, and interpretation activities.

The first clue to the possible seriousness of attrition arose after reminders were mailed to participants to encourage them to pretest their pupils. It was apparent that many subjects refused to pretest due to preconceived notions about the appropriateness of the test for their pupils. Many thought the test might have been too difficult.

While no scoring keys were provided the 290 music educators for the music achievement tests, it was evident that many subjects declined to posttest their pupils after seeing their pretest data. This was, perhaps due to a misunderstanding about testing techniques and a general feeling that their pupils had scored poorly on the pretest and, therefore, would not do well on the posttest thereby affecting the music educator's chances for being considered

BEST COPY AVAILABLE

a successful music teacher. Most of these subjects did not realize that a low pupil score on an appropriate and valid music achievement pretest was, in fact, desirable because it allowed more room for growth even though these regression effects were, to some extent, controlled statistically.

Most of the music educators who failed to posttest their pupils did not return the testing packages to the project office even though reminders were mailed them, and the investigator and his staff telephoned these delinquent subjects. In only 5 cases did the posttesting packets arrive at the music educator's school too late for administering.

More subtle kinds of subject attrition occurred due to posttesting errors, and these were more difficult to interpret. While it was made perfectly clear to the participating music educators that one class of pupils was to be used for the music achievement testing (and this class preferably would not be their most musically sophisticated group of pupils), some subjects tested more than one class.

There were numerous problems in matching pre and posttests for certain classes of students. This produced a possible pupil testing bias; and, therefore, if less than 13 pupils could be matched from one music class, the data were not used. For example, two music educators pretested one class and posttested a different class of pupils. Furthermore, one music educator pretested his pupils with one music achievement test and posttested with a different test. In this case, the music educator selected the complete Music Achievement Test III and was mailed the record and the machine scored answer sheets. He used "Side One" for the pretest and "Side Two" for the posttest.

Listed below are the findings for the attrition from 290 to 209 subjects:

REASONS	NUMBER OF SUBJECTS
1. Music achievement tested another music teacher's class	1
2. Music achievement tested different sections for pre and posttests	1
3. Music achievement tested one class for pretesting and a different class for posttesting	2
4. Attempted to sabotage the testing by testing a number of pupils of whom less than 5 matched for pre and posttests	3
5. Music achievement tested fewer than 13 pupils	22
6. Failed to pre and posttest their pupils	19
7. Failed to posttest their pupils	20
8. Did not have 3 or more rating evaluations	7
9. Failed to undertake any criterion efforts	5
10. Death	1
Total	81

There was no sufficient way in which to gather evidence as to why the attrition occurred other than through information provided the investigator in verbal and written communications. Perhaps it was best summarized by the letter received from the research director for a large public school system in the midwest who approved the study:

> I should point out that the vote on acceptance of your research was close. The major criticism seemed to focus on the large amount of student and teacher time that would be involved in the conduct of your study. This is indeed an important factor. I minimize this fact, however, in light of the fact that teacher participation in all research studies is optional as far as they are concerned.
>
> You should be aware, therefore, that while approval to conduct your research is given, that this in no way guarantees teacher or pupil involvement. Their willingness to engage in your study will depend obviously on what value they see in the potential results. I am sure our music supervisor will be helpful to you in explaining the merits and the need for the study to our music teachers (personal communication to the investigator, June 23, 1971).

The above advice was followed in that the staff of this school district took considerable time in approving the study. Since this city was one of the first large school districts to participate (and others would be invited), it was deemed appropriate to provide ways in which the study could be explained to prospective subjects, allowing them the opportunity to volunteer or drop out as they valued the study in terms of their and their pupils' time commitments.

The music educator sample (N=209) was stratified into four subsamples according to teacher type and sex: female choral, female instrumental, male choral, and male instrumental. Table 4 shows the attrition of these subsamples from the original 290 subjects to the final 209 participants.

TABLE 4.--MUSIC EDUCATOR ATTRITION

	In	Dropouts	Total	Percentage of original sample	Percentage of final sample
Female Choral	103	29	132	45	49
Female Instrumental	12	11	23	8	6
Male Choral	26	11	37	13	12
Male Instrumental	68	30	98	34	33
Totals	209	81	290	100	100

The sample of 290 music educators was weighted in favor of the female choral and male instrumental music teachers. It was almost evenly divided between male and female subjects, however, with 53 percent female and 47 percent male. Attrition was most severe in the female instrumental group where just under 50 percent dropped out of the study. In terms of the total sample, the final sample percentages were quite close to those originally obtained: a change from 45 percent to 49 percent for female choral, from 8 percent to 6 percent for female instrumental, from 13 percent to 12 percent for male choral, and from 34 percent to 33 percent for male instrumental. The sample appeared to remain relatively stable within subgroups even though total attrition accounted for 28 percent of the original sample of 290 subjects.

As was discussed in Chapter I, a gratuity was paid to certain music educators teaching in large school districts in order to obtain their cooperation and to reduce sample volunteer bias effects. Table 5 highlights the music educator attrition rate within subgroups comparing paid and unpaid music educators.

The total paid music educators represented 57 percent of the sample before attrition while 43 percent were unpaid. After attrition, 55 percent of the sample were paid while 45 percent were unpaid. The sample remained relatively

TABLE 5.--MUSIC EDUCATOR ATTRITION - PAID AND UNPAID

	Paid	In	Out	Follow Through Percentage	Unpaid	In	Out	Follow Through Percentage
FC	86	64	22	74	46	39	7	85
FI	12	7	5	58	11	5	6	45
MC	16	14	2	88	21	12	9	57
MI	52	30	22	58	46	28	8	83
Totals	166	115	51	69	124	94	30	76

Code: FC=Female Choral, FI=Female Instrumental,
MC=Male Chorak, MI=Male Instrumental

stable within subgroups percentage wise and within the paid and unpaid groups. However, one very interesting factor occurred. The follow through percentage for the groups with the largest number of subjects (female choral and male instrumental) were high for the unpaid subjects while the subjects in these two groups who were paid followed through with their obligations at a lower percentage rate. Interesting though, the two smallest groups (female instrumental and male choral) had reverse follow through percentages - the paid groups followed through more with their obligations than did the volunteer unpaid subjects. These two subgroups helped to maintain sample attrition at a relatively low rate. Table 6 shows the percentage after attrition for the music educator sample.

TABLE 6.--MUSIC EDUCATOR SAMPLE AFTER ATTRITION (N=209)

	Paid Percentage	Unpaid Percentage	Total Sample Percentage
Female Choral	56	42	49
Female Instrumental	6	5	6
Male Choral	12	13	12
Male Instrumental	26	40	33
Totals	100	100	100

Table 7 indicates that some of the objectives were achieved by offering the gratuity and paying half the sample. More subjects participated who were teaching in combination schools (17 percent to 11 percent) and disadvantaged schools (31 percent to 13 percent). While still weighted in favor of the advantaged schools, the final sample (N=209) included 63 percent advantaged, 15 percent combination, and 22 percent disadvantaged. Without the addition of

TABLE 7.--MUSIC EDUCATOR SOCIOECONOMIC SCHOOL CATEGORY DISTRIBUTION

	ADVANTAGED				COMBINATION				DISADVANTAGED				TOTALS			
	IN		OUT		IN		OUT		IN		OUT		IN		OUT	
PAID SUBJECTS	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
Female Choral	33	29	07	14	09	08	04	08	22	19	11	21	64	56	22	43
Female Instrumental	05	04	03	06	01	01	01	02	01	01	01	02	07	06	05	10
Male Choral	07	06	00	00	05	04	01	02	02	02	01	02	14	12	02	04
Male Instrumental	15	13	11	21	05	04	06	12	10	09	05	10	30	26	22	43
Total Paid	60	52	21	41	20	17	12	24	35	31	18	35	115	100	51	100
UNPAID SUBJECTS																
Female Choral	30	32	05	167	05	05	00	00	04	043	02	066	39	413	07	233
Female Instrumental	02	02	04	134	01	01	00	00	02	020	02	066	05	050	06	200
Male Choral	09	10	08	267	01	01	00	00	02	020	01	034	02	130	09	301
Male Instrumental	30	32	06	200	04	043	00	00	04	043	02	066	38	406	08	266
TOTAL UNPAID	71	76	23	768	11	113	00	00	12	126	07	232	94	999	30	1000
TOTAL SUBJECTS																
Female Choral	63	301	12	148	14	067	04	049	26	124	13	160	103	492	29	357
Female Instrumental	07	034	07	087	02	009	01	012	03	014	03	037	12	057	11	136
Male Choral	16	077	08	099	06	029	01	012	04	019	02	025	26	125	11	136
Male Instrumental	45	215	17	210	09	043	06	074	14	067	07	087	68	325	30	371
Total Sample	131	627	44	544	31	148	12	147	47	224	25	309	209	999	81	1000

the paid subjects, the respective figures would have been 76 percent advantaged, 11 percent combination, and 13 percent disadvantaged.

Table 7 shows that the within subjects, paid-unpaid, and socioeconomic school classification groups remained stable as far as attrition was concerned. While the sample was weighted in favor of the advantaged, male instrumental, and female choral groups, attrition did not contribute to these weightings.

An additional factor was that music programs varied widely from district to district and school to school within the socioeconomic school classifications, and therefore, the situation was accounted for in the treatment of the gain score data. The following tables compare the rates of attrition for the sample in terms of music teacher type, the socioeconomic category, district size, and the paid and unpaid subjects' mean ages, years of teaching experience, and degrees. The degree code represents the following numerical breakdown:

1 = bachelor's degree
2 = bachelor's degree plus 12 to 15 credits
3 = bachelor's degree plus 16 to 30 credits
4 = master's degree
5 = master's degree plus 24 credits
6 = master's degree plus 36 credits (Advanced Certificate)
7 = master's degree plus 64 credits (ABD degree - all but dissertation)
8 = doctor's degree

These tables are presented to the reader for attrition purposes only. The subjects were not stratified into all these subgroups for personality and motivational analyses.

TABLE 8.--FEMALE CHORAL IN STUDY

School Classification	N	Age	Years of Teaching Experience	Degree
			PAID TEACHERS - LARGE DISTRICT	
Advantaged	33	34.88	9.46	2.88
Combination	9	35.78	9.44	2.00
Disadvantaged	21	36.36	10.91	2.50
Total	63	35.67	9.94	2.46
			UNPAID TEACHERS - LARGE DISTRICT	
Advantaged	19	32.90	7.68	2.42
Combination	3	28.00	6.67	3.33
Disadvantaged	4	42.33	5.00	3.00
Total	26	34.41	6.45	2.92
			UNPAID TEACHERS - SMALL DISTRICT	
Advantaged	11	33.09	7.36	1.91
Combination	2	26.00	4.00	1.50
Disadvantaged	1	35.00	12.00	3.00
Total	14	31.36	7.79	2.14
			TOTAL FEMALE CHORAL TEACHERS	
Advantaged	63	33.62	8.17	2.40
Combination	14	29.93	6.70	2.28
Disadvantaged	26	37.90	9.30	2.83
Mean Total of School Classifications	103	33.82	8.06	2.51
Mean Total of Subtotals	103	33.82	8.06	2.51

TABLE 9.--FEMALE CHORAL OUT OF STUDY

School Classification	N	Age	Years of Teaching Experience	Degree
			PAID TEACHERS - LARGE DISTRICT	
Advantaged	7	39.71	10.71	2.14
Combination	4	35.50	8.75	3.75
Disadvantaged	11	32.73	8.64	2.55
Total	22	35.98	9.37	2.81
			UNPAID TEACHERS - LARGE DISTRICT	
Advantaged	2	23.50	1.50	1.50
Combination	0			
Disadvantaged	1	30.00	2.00	1.00
Total	3	26.75	1.75	1.25
			UNPAID TEACHERS - SMALL DISTRICT	
Advantaged	3	30.00	4.33	1.00
Combination	0			
Disadvantaged	1	40.00	8.00	4.00
Total	4	35.00	6.17	2.50
			TOTAL FEMALE CHORAL OUT OF STUDY	
Advantaged	12	31.07	5.52	1.55
Combination	4	35.50	8.75	3.75
Disadvantaged	13	34.24	6.21	2.52
Mean Total of School Classifications	29	33.60	6.83	2.60
Mean Total of Totals	29	32.58	5.76	2.19

TABLE 10.--FEMALE INSTRUMENTAL IN STUDY

School Classification	N	Age	Years of Teaching Experience	Degree
			PAID TEACHERS - LARGE DISTRICT	
Advantaged	5	28.60	7.00	2.20
Combination	1	37.00	14.00	4.00
Disadvantaged	1	28.00	6.00	1.00
Total	7	31.20	9.00	2.40
			UNPAID TEACHERS - LARGE DISTRICT	
Advantaged	1	25.00	4.00	2.00
Combination	1	27.00	5.00	1.00
Disadvantaged	1	26.00	2.00	4.00
Total	3	26.00	3.67	2.33
			UNPAID TEACHERS - SMALL DISTRICT	
Advantaged	1	26.00	5.00	4.00
Combination	0			
Disadvantaged	1	23.00	1.00	1.00
Total	2	24.50	3.00	2.50
			TOTAL FEMALE INSTRUMENTAL IN STUDY	
Advantaged	7	26.53	5.33	2.73
Combination	2	32.00	9.50	2.50
Disadvantaged	3	25.67	3.00	2.00
Mean Total of School Classifications	12	28.07	5.94	2.41
Mean Total of Totals	12	27.23	5.33	2.41

TABLE 11.--FEMALE INSTRUMENTAL OUT OF STUDY

School Classification	N	Age	Years of Teaching Experience	Degree
			PAID TEACHERS - LARGE DISTRICT	
Advantaged	3	30.33	5.33	3.00
Combination	1	51.00	7.00	3.00
Disadvantaged	1	37.00	15.00	5.00
Total	5	39.44	9.11	3.67
			UNPAID TEACHERS - LARGE DISTRICT	
Advantaged	2	47.00	22.00	3.50
Combination				
Disadvantaged	1	22.00	1.00	1.00
Total	3	34.50	11.50	2.25
			UNPAID TEACHERS - SMALL DISTRICT	
Advantaged	2	41.50	12.50	3.50
Combination	0			
Disadvantaged	1	23.00	2.00	2.00
Total	3	32.25	7.25	2.75
			TOTAL FEMALE INSTRUMENTAL OUT OF STUDY	
Advantaged	7	39.61	13.28	3.33
Combination	1	51.00	7.00	3.00
Disadvantaged	3	27.33	6.00	2.67
Mean Total of School Classifications	11	39.32	8.76	3.00
Mean Total of Totals	11	35.40	9.29	2.90

TABLE 12.--MALE CHORAL IN STUDY

School Classification	N	Age	Years of Teaching Experience	Degree
			PAID TEACHERS - LARGE DISTRICT	
Advantaged	7	34.86	10.57	4.29
Combination	5	45.60	20.60	2.20
Disadvantaged	2	37.00	9.00	3.00
Total	14	39.15	13.39	3.16
			UNPAID TEACHERS - LARGE DISTRICT	
Advantaged	5	36.00	12.60	4.60
Combination	1	33.00	11.00	5.00
Disadvantaged	1	41.00	17.00	6.00
Total	7	36.67	13.53	5.20
			UNPAID TEACHERS - SMALL DISTRICT	
Advantaged	4	31.50	3.00	2.75
Combination	0			
Disadvantaged	1	33.00	8.00	7.00
Total	5	32.25	5.50	4.88
			TOTAL MALE CHORAL IN STUDY	
Advantaged	16	34.12	8.72	3.79
Combination	6	39.30	15.80	3.60
Disadvantaged	4	37.00	11.33	5.33
Mean Total of School Classifications	26	36.81	11.95	4.24
Mean Total of Totals	26	36.02	10.81	4.41

TABLE 13.--MALE CHORAL OUT OF STUDY

School Classification	N	Age	Years of Teaching Experience	Degree
			PAID TEACHERS - LARGE DISTRICT	
Advantaged	0			
Combination	1	30.00	5.00	2.00
Disadvantaged	1	33.00	7.00	3.00
Total	2	31.50	6.00	2.50
			UNPAID TEACHERS - LARGE DISTRICT	
Advantaged	4	36.00	11.75	3.75
Combination	0			
Disadvantaged	1	24.00	2.00	1.00
Total	5	30.00	6.86	2.38
			UNPAID TEACHERS - SMALL DISTRICT	
Advantaged	4	29.25	7.00	3.00
Combination	0			
Disadvantaged	0			
Total	4	29.25	7.00	3.00
			TOTAL MALE CHORAL OUT OF STUDY	
Advantaged	8	32.63	9.38	3.38
Combination	1	30.00	5.00	2.00
Disadvantaged	2	28.50	4.50	2.00
Mean Total of School Classifications	11	30.38	6.29	2.46
Mean Total of Totals	11	30.25	6.63	2.63

TABLE 14.--MALE INSTRUMENTAL IN THE STUDY

School Classification	N	Age	Years of Teaching Experience	Degree
			PAID TEACHERS - LARGE DISTRICT	
Advantaged	15	40.40	14.20	3.68
Combination	5	40.60	13.80	3.00
Disadvantaged	10	41.30	14.90	2.70
Total	30	40.77	14.30	3.12
			UNPAID TEACHERS - LARGE DISTRICT	
Advantaged	20	31.55	9.55	3.65
Combination	2	41.50	19.00	5.00
Disadvantaged	1	26.00	5.00	2.00
Total	23	33.01	11.18	3.55
			UNPAID TEACHERS - SMALL DISTRICT	
Advantaged	10	35.10	11.80	4.10
Combination	2	25.00	2.00	2.50
Disadvantaged	3	36.67	12.33	5.00
Total	15	32.26	8.71	3.87
			TOTAL MALE INSTRUMENTAL IN STUDY	
Advantaged	45	35.68	11.85	3.81
Combination	9	35.70	11.60	3.50
Disadvantaged	14	34.66	10.74	3.23
Mean Total of School Classifications	68	35.35	11.40	3.51
Mean Total of Totals	68	35.35	11.40	3.51

TABLE 15.--MALE INSTRUMENTAL OUT OF STUDY

School Classification	N	Age	Years of Teaching Experience	Degree
			PAID TEACHERS - LARGE DISTRICT	
Advantaged	11	40.00	15.70	3.70
Combination	6	34.68	8.33	2.83
Disadvantaged	5	32.00	7.40	2.60
Total	22	35.56	10.48	3.04
			UNPAID TEACHERS - LARGE DISTRICT	
Advantaged	2	43.00	19.00	3.50
Combination	0			
Disadvantaged	2	41.00	18.50	4.00
Total	4	42.00	18.75	3.75
			UNPAID TEACHERS - SMALL DISTRICT	
Advantaged	4	32.75	7.75	3.25
Combination	0			
Disadvantaged	0			
Total	4	32.75	7.75	3.25
			TOTAL MALE INSTRUMENTAL OUT OF STUDY	
Advantaged	17	38.58	14.15	3.83
Combination	6	34.68	8.33	2.83
Disadvantaged	7	36.50	12.95	3.30
Mean Total of School Classification	30	36.58	11.81	3.32
Mean Total of Totals	30	36.77	12.33	3.35

BEST COPY AVAILABLE

Music Student Teacher Sample

The music student teacher sample was selected primarily from volunteer music education majors attending 7 teacher training institutions in the following categories: medium sized private universities (Elmhurst College, Millikin University, and Oklahoma City University); medium sized public universities (Eastern Illinois University and Western Illinois University); and large public universities (Illinois State University and the University of Illinois). The first semester, 1971-72 school year, student teachers attending the University of Illinois and the first semester 1972-73 school year student teachers attending Oklahoma City University were required to participate in the study thus reducing volunteer bias.

From a population of 183 music student teachers, a total of 93 senior music education majors completed the 3 personality tests. The attrition from 183 potential subjects to 93 participants represented 51 percent of the originally contacted students. This response was considered adequate and within the realms of normality as far as attrition was concerned.

Similar to the music educator sample, a second kind of attrition occurred which reduced the sample from 93 to 73 subjects. This represented a 22 percent attrition rate.

Table 16 shows the distribution of the music student teacher sample in 3 school and 4 teacher type categories.

TABLE 16.--MUSIC STUDENT TEACHER SAMPLE - NUMBER OF SUBJECTS

	Medium Private			Medium Public			Large Public			Total Sample		
	In	Out	Total	In	Out	Total	In	Out	Total	In	Out	Total
Female Choral	1	1	2	7	3	10	24	5	29	32	9	41
Female Instr.	2	0	2	3	0	3	7	1	8	12	1	13
Male Choral	0	0	0	1	1	2	3	1	4	4	2	6
Male Instr.	5	1	6	7	2	9	13	5	18	25	8	33
Total	8	2	10	18	6	24	47	12	59	73	20	93

The following are the percentages of the original to the final sample for the music student teachers by teacher training institution and teacher type:

TABLE 17.--STUDENT TEACHER SAMPLE PERCENTAGE FOR ATTRITION

		Original	Final
Medium Private School		11%	11%
Medium Public School		26%	25%
Large Public School		63%	64%
	Total	100%	100%
Female Choral Majors		44%	44%
Female Instrumental Majors		14%	16%
Male Choral Majors		7%	6%
Male Instrumental Majors		35%	34%
	Total	100%	100%

Attrition did not affect the weightings of the original to the final sample in terms of numbers of music student teachers represented by the size of the teaching training institution or by the music major type groups. The sample was weighted in favor of the large public university group and the female choral and male instrumental teacher type groups. Attrition did not affect the distribution of the subsamples.

THE PUPIL GROWTH CRITERION OF MUSIC TEACHING SUCCESS

Subtests from the Music Achievement Tests I-IV by Richard J. Colwell were used to compute the class mean residual gain score for each participating music educator and special music achievement tests were prepared for each subject's class. This provided a criterion referenced approach to the music achievement testing because the music educator selected the objectives which he considered important and for which he taught. Music teachers teaching at the high school level were instructed, if possible, to use a less skilled group of pupils for music achievement testing rather than use their best performing group of pupils.

Scoring

Hand scoring accuracy was checked by the project consultant who randomly selected 41 test answer sheets from the 23, 596 available. Following is the formula:

$$\text{percent } \frac{1}{N} = f \times 100$$

One error was found in scoring. This error can be interpreted according to the above equation in three ways. First, the total number of test items on the 41 answer sheets was 2,261 of which one was in error. The scoring accuracy percentage of this is 99.96 percent.

Second, the total number of subtests included on the answer sheets was 123 of which 122 were scored correctly. This produced a scoring accuracy figure of 99.19 percent.

Third, the total number of test answer sheets numbered 41 of which 40 were scored correctly. This scoring accuracy figure was 97.56 percent.

The mean scoring accuracy percentage which was found by adding the three percentages and dividing by 3 produced a figure of 98.9 percent. Rounding produced a 99 percent accuracy for scoring the music achievement tests by hand. The 1 percent error figure was substantially lower than the 28 percent average figure found for hand scoring standardized tests:

> Accuracy cannot be overemphasized in the scoring of standardized tests. Even small scoring errors have a detrimental influence on the reliability and validity of the test results. Large errors may lead to educational decisions which are injurious to individuals or groups (Gronlund 1971:353).

Residual Gain Scores

A multiple regression equation which included certain organismic, demographic, and experiential variables was used in order to derive the class mean residual gain score for each participating music educator. The literature discussed in Chapter II was inconclusive as far as the specific technique or formula to be used. While it had been hypothesized that variables other than the pretest score should be used as covariates, little research had been offered to assist this investigation with the specific procedural requirements for formulating the residual gain equation. Therefore, the investigator selected the residual technique which appeared to control best the pre-existing differences among teachers' classes of pupils. Table 18 shows the class mean residual gain score analysis.

TABLE 18.--CLASS MEAN RESIDUAL GAIN SCORE MULTIPLE REGRESSION ANALYSIS

	β COEFFICIENTS	b COEFFICIENTS	SE of β	SE of b	t-test
	Y	Y	Y	Y	Y
Intercept				1.332	6.093**
X_1	.015	.004	.036	.010	.416
X_2	-.007	-.084	.035	.395	-.212
X_3	.925	.912	.027	.027	34.258**

Intercept 8.118** N = 211 DF = 3/207

F = 432.38 R = .93 $R^2 = .86$ SE of Estimate = 3.88 $p<.01$**

The finding from the multiple regression equation used to obtain the class mean residual gain score was that the created variable (X_1) and the meeting times per week variable (X_2) did not substantially contribute to the class mean residual gain score. Therefore, the residual gain score analysis appeared to be quite similar to analysis of covariance in that the class mean pretest score contributed significantly the most to the equation and the derivation of the residual gain score. The F ratio indicated that the multiple R was significant beyond the .01 level and that 86 percent of the variance was accounted for by the equation in formulating the class mean residual gain score for each participating music educator, the G criterion.

After computing the G criterion score for each participant, it was necessary to investigate possible differences in G which could be attributed to pretesting time differences. Analysis of variance and t-test procedures were used to test for significant differences among these pretesting time groups. The following null hypothesis was investigated:

> There will be no significant differences among the four pretesting time groups in class mean residual gain scores (G).

Table 19 shows the results of the analysis of variance procedure.

TABLE 19.--ANOVA CLASS MEAN RESIDUAL GAIN SCORES

Pretesting Times	Number	Mean	Standard Deviation
1 = September-October	75	.901	3.491
2 = November-December	63	.136	4.777
3 = January-February	66	-.773	2.391
4 = March	7	-3.594	6.260

	Degrees of Freedom	Sum of Squares	Mean Squares	F Ratio
Source Between	3	191.945	63.981	4.53**
Within	207	2923.709	14.124	
	210	3115.654		

p<.01**

The analysis of variance showed significant differences among the groups beyond the .01 level for a two tailed test of significance. Because significant differences were found to exist, t-test analyses were used to investigate where these differences were located. Table 20 shows the t-test analysis for testing the null hypotheses that there would be no significant differences between the pretesting groups.

TABLE 20.--t-TEST ANALYSIS AMONG PRETESTING GROUPS

Groups	DF	Mean	N	Difference Between Means	Pooled SE	t-Test Values
1 & 2	136	.552	138	.765	.705	1.085
1 & 3	139	.117	141	1.674	.511	3.277**
1 & 4	81	.505	83	4.107	1.399	2.936**
2 & 3	127	-.329	129	.909	.661	1.376
2 & 4	69	-.241	71	3.342	1.840	1.816
3 & 4	72	-1.036	74	2.433	1.094	2.223*
1-2, 3-4	210	-.003	212	1.588	.544	2.919**
1-3, 2-4	210	-.003	212	.358	.560	.639

p<.05* p<.01**

Table 20 shows that there were significant differences between groups 1 and 3, groups 1 and 4, and groups 1-2 and 3-4 beyond the .01 level. There was also a significant difference between groups 3 and 4 beyond the .05 level. Therefore, significant differences in class mean residual gain scores which could be attributed to pretesting time differences were found for groups 1, 3, and 4. Even though there were no significant differences between groups 2 and 3 and groups 2 and 4, when the first two groups were combined into one group and compared with a group including both groups 3 and 4, significant differences were found. Furthermore, because a significant difference was found between group 1 and group 3-4, the music educator sample was stratified into three pretesting groups: Group 1-2 reflecting pretesting times during the first semester months of September, October, November, and December; Group 3 reflecting pretesting during the first two months of the second semester (January and February); and Group 4 reflecting pretesting during March.

Music educators were rank ordered from high to low according to class mean residual gain scores within each of the three pretesting groups and within subgroups of teacher type: female choral, female instrumental, male choral, and male instrumental. The total group stratification was accomplished by percentage allocation according to the number of subjects in the pretesting time groups and in the within teacher type groups. This procedure allowed for a total G criterion ranked order for the complete sample of music educators (N=[illegible]). In other words, the subjects from each of the three pretesting time groups were given a rank according to the percentage of subjects in each pretesting, teacher type group.

Table 21 shows the music achievement test subtests which were used to compute the G criterion. Tables 18 through 20 show the complete G criterion scores for the music educator samples within subgroups according to teacher type.

TABLE 21.--MUSIC ACHIEVEMENT TEST SUBTESTS

Test B	Pitch Discrimination - To differentiate which tone of 2 musical tones is higher or whether the tones are the same and to differentiate which of 3 musical tones heard is the lowest. MAT-I, Part 1.
Test C	Interval Discrimination - To differentiate scalewise from leaping intervals. MAT-I, Part 2.
Test D	Meter Discrimination - To differentiate duple and triple meters; music that moves in twos from music that moves in threes. MAT-I, Part 3.
Test E	Major-Minor Mode Discrimination - To distinguish major and minor modes in music. MAT-II, Part 1.
Test F	Feeling For Tonal Center - To determine the key center or key tone of a group of chords in one tonality and of a short musical phrase. MAT-II, Part 2.
Test H	Auditory-Visual Discrimination: Pitch - To read notes accurately as to pitch. MAT-II, Part 3A.
Test I	Auditory-Visual Discrimination: Rhythm - To read notes accurately as to rhythm. MAT-II, Part 3B.
Test J	Tonal Memory - To recognize whether two chords are the same or different and where different, e.g. to find which note was altered. MAT-III, Part 1.
Test K	Melody Recognition - To recognize the melody and determine whether it appears in the top, middle, or bottom voice of a harmonized version. MAT-III, Part 2.
Test L	Instrument Recognition - To identify the instruments of the band and orchestra by sound. MAT-III, Part 4.
Test M	Musical Style: Composers - To identify music of different styles by selecting representative composers. MAT-IV, Part 1A.
Test N	Musical Style: Texture - To differentiate monophonic, homophonic, and polyphonic music. MAT-IV, Part 1B.
Test O	Cadence Recognition - To differentiate half, full and deceptive cadences. MAT-IV, Part 4.

MUSIC ACHIEVEMENT TESTS PREPARED FROM ABOVE SUBTESTS

01	BCD	11	BEL	21	DIK	31	HIJ	41	BLM
02	BCL	12	DFI	22	DIL	32	HIK	42	MAT-III
03	BDE	13	DHL	23	EHJ	33	HIL	43	MAT-II (HIEF)
04	BDF	14	BIJ	24	EIK	34	HKL	44	LMNO
05	BDI	15	BIL	25	EIL	35	HKN	45	MAT-IV
06	BDJ	16	BJL	26	EJK	36	JMN	46	BD
07	BDL	17	DEF	27	HIB	37	KIL	47	BCDE
08	BDN	18	DEI	28	HID	38	KIM		
09	BEI	19	DEK	29	HIE	39	KLM		
10	BEK	20	DEL	30	HIF	40	BD½F		

The interpretation of the Codes for Tables 22 through 25 follows:

Grade Level - H = high school, grades 9-12
M = middle school, grades 6-9
E = elementary school, grades 5-6

Pupil N - The number of pupils used for pre and posttesting with the specially prepared music achievement tests.

School Code - The socioeconomic nature of the school:
A = socially advantaged
C = combination
D = socially disadvantaged

School District - The category for the school district:
LP = large district, paid subjects
LO = large district, unpaid subjects
SO = small district, unpaid subjects

MAT Test - The number of the specially prepared music achievement test used for pre and posttesting Table 21.

Highest Score - The highest possible score for that specially prepared music achievement test.

Pretest Mean - The pretest class mean score on the specially prepared music achievement test.

Pretest SD - The pretest standard deviation score for the class mean pretest score.

Posttest Mean - The class mean posttest score on the specially prepared music achievement test.

Posttest SD - The standard deviation for the class mean posttest score.

Raw Gain - The difference found by subtracting the pretest class mean score from the class mean posttest score, not statistically treated.

Res. Gain - The residual gain score. The statistically treated class mean residual gain score which was a deviation score derived from the multiple regression equation which included 3 covariates and predicted a posttest class mean score. This predicted score was subtracted from the actual class mean posttest score which derived the class mean residual gain score for the G criterion.

Pretest Group - 1 = September, October, 2 = November, December
3 = January, February 4 = March

Gain Curve - The gain success score which was assigned to each music educator within each subgroup and reflected the normal curve distribution of his residual gain G score. The scores ranged from a high of 6 to a low of 1 and reflected 3 standard deviations above and below the mean rank score for the sample.

6 = outstanding, 3 percent of the sample, 3 standard deviations above the mean rank.
5 = excellent, 13 percent of the sample, 2 standard deviations above the mean rank.
4 = above average, 34 percent of the sample, 1 standard deviation above the mean rank.
3 = below average, 34 percent of the sample, 1 standard deviation below the mean rank.
2 = poor, 13 percent of the sample, 2 standard deviations below the mean rank.
1 = unsuccessful, 3 percent of the sample, 3 standard deviations below the mean rank.

Interval Pre-Post - The actual numerical interval for instruction between the administrations of the pre and posttests.

Rank Num. - The rank number assigned each subject within the three pretesting time groups.

TABLE 22.--FEMALE CHORAL GAIN SUCCESS

Grade Level	Pupil N	Schl Code	Schl Dist	MAT Test	High Score	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Raw Gain	Res. Gain	Pretest Group	Gain Curve	Interval Pre-Post	Rank Num.
E	21	A	SO	07	70	39.14	7.01	58.86	7.17	19.72	14.97	1	6	31	2
E	24	C	LO	13	68	37.92	11.60	52.00	8.42	14.08	9.30	2	6	23	4
E	26	D	LP	47	111	53.08	10.82	63.42	11.35	10.34	6.95	1	5	18	7
M	34	A	SO	29	88	37.38	13.61	48.97	14.39	11.59	6.52	1	5	21	8
E	27	A	LP	47	111	53.96	9.64	62.74	10.15	8.78	5.40	1	5	18	11
E	18	A	LP	47	111	58.83	7.95	66.78	8.80	7.95	5.02	1	5	18	12
E	14	A	LP	47	111	52.86	6.98	60.86	9.55	8.00	4.54	1	5	18	15
E	28	A	SO	07	70	42.14	9.69	50.96	8.30	8.82	4.36	1	5	20	17
E	25	A	LO	01	83	[illegible]	[illegible]	56.36	12.08	8.12	4.23	1	5	21	18
E	20	C	LO	01	83	46.20	[illegible]	54.20	13.09	8.00	3.98	1	5	18	21
H	17	A	LO	01	83	47.71	[illegible]	55.47	10.87	7.76	3.84	1	4	19	22
E	27	A	SO	32	80	29.56	12.21	38.89	12.48	9.33	3.80	1	4	18	24
E	24	A	SO	03	83	43.13	7.84	50.63	8.88	7.50	3.18	1	4	19	27
E	19	A	LP	47	111	59.42	10.36	65.47	14.10	6.05	3.17	1	4	18	28
E	38	A	LO	07	70	43.39	9.45	50.84	10.07	7.45	3.11	1	4	18	29
E	15	A	LP	33	75	21.07	5.81	30.33	8.83	9.26	2.80	2	4	16	31
E	23	A	LO	01	83	44.96	8.11	51.78	12.30	6.82	2.64	1	4	21	32
E	16	C	LO	17	70	32.81	6.96	40.50	9.91	7.69	2.47	1	4	22	34
E	19	A	SO	27	85	35.21	11.58	42.68	10.75	7.47	2.37	2	4	31	35
M	22	D	LP	07	70	29.82	7.52	37.23	8.34	7.41	1.91	1	4	19	39
E	30	A	SO	03	83	41.00	11.82	47.07	12.47	6.07	1.58	1	4	20	43
E	20	A	LO	05	87	36.75	11.74	43.05	12.19	6.30	1.41	2	4	18	44
E	24	D	LP	47	111	53.54	6.34	58.13	7.56	4.59	1.24	1	4	18	47
E	21	C	LP	47	111	56.90	10.73	60.95	10.34	4.05	.97	1	4	18	49
E	30	A	LP	47	111	65.57	13.26	68.87	11.55	3.30	.96	1	4	18	50
E	26	C	LP	47	111	57.58	11.51	61.54	14.40	3.96	.94	1	4	18	51
E	23	A	LO	05	87	32.91	9.39	38.96	9.91	6.05	.85	1	4	19	53
E	21	A	LO	01	83	53.52	9.26	57.38	9.59	3.86	.47	2	4	17	57
E	28	A	LO	28	90	30.29	10.67	36.14	12.34	5.85	.41	1	4	20	58
H	67	C	SO	32	80	44.04	16.41	48.70	14.54	4.66	.39	1	4	16	59
M	52	A	LO	11	68	38.90	9.96	43.88	10.04	4.98	.27	2	4	10	63
E	23	A	LO	05	87	39.61	12.35	44.39	11.08	4.78	.16	2	4	22	65
H	21	A	SO	29	88	47.38	14.18	51.43	21.96	4.05	- .10	1	4	19	67
E	23	D	LP	03	83	33.04	9.40	38.04	11.31	5.00	- .14	2	3	22	68
H	23	A	LO	07	70	46.87	10.71	50.70	10.76	3.83	- .21	2	3	19	69
E	19	D	LP	47	111	52.42	10.20	55.63	9.39	3.21	- .24	1	3	18	71
E	17	A	LO	01	83	52.18	11.71	55.35	11.09	3.17	- .35	1	3	19	73

BEST COPY AVAILABLE

TABLE 22.--CONTINUED

Grade Level	Pupil N	Schl Code	Schl Dist	MAT Test	High Score	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Raw Gain	Res. Gain	Pretest Group	Gain Curve	Interval Pre-Post	Rank Num.
E	33	D	LO	03	83	40.82	8.26	44.88	6.91	4.06	- .41	1	3	16	75
E	29	C	LP	27	85	29.52	7.87	34.24	9.85	4.72	- .51	2	3	20	77
E	24	C	LP	47	111	61.54	11.13	63.54	12.28	2.00	- .68	1	3	18	80
E	17	D	LO	01	83	38.24	11.90	42.00	11.44	3.76	- .88	2	3	19	84
E	37	A	LP	02	68	40.22	10.05	43.78	11.80	3.56	-1.03	2	3	17	88
E	27	A	LO	01	83	52.33	12.67	54.78	15.09	2.45	-1.07	1	3	20	89
E	29	A	LO	03	83	38.38	10.09	42.00	8.67	3.62	-1.13	1	3	18	90
E	23	A	LO	07	70	30.78	7.26	35.04	6.91	4.26	-1.14	2	3	17	91
E	29	D	LP	27	85	27.90	9.40	32.21	11.01	4.31	-1.29	1	3	23	95
E	26	A	LO	01	83	40.62	9.64	43.85	11.52	3.23	-1.30	2	3	19	96
H	126	A	LP	32	80	53.02	16.21	55.46	15.56	2.44	-1.33	2	3	22	97
E	32	A	LO	01	83	41.19	12.58	44.31	12.67	3.12	-1.37	1	3	19	100
M	33	A	LP	27	85	53.39	16.49	55.33	15.71	1.94	-1.82	2	3	23	104
M	23	A	LO	03	83	50.61	8.41	52.35	9.41	1.74	-1.97	1	3	25	105
M	24	D	LP	19	78	37.17	8.85	39.96	8.04	2.79	-2.01	2	3	18	106
E	24	D	LP	47	111	52.67	9.89	54.08	10.78	1.41	-2.02	1	3	18	107
E	27	A	LO	05	87	38.63	12.12	41.11	13.97	2.48	-2.22	2	3	18	109
E	28	D	LP	15	72	20.21	6.71	24.36	8.25	4.15	-2.23	1	3	33	110
E	23	D	LP	03	83	35.43	6.97	38.00	12.83	2.57	-2.30	1	3	20	111
E	25	A	LO	03	83	49.64	10.80	50.76	12.27	1.12	-2.65	2	3	20	112
E	19	A	LP	33	75	27.26	9.13	30.58	9.30	3.32	-2.80	2	2	23	114
E	14	D	LO	28	90	38.70	5.77	40.37	3.91	1.67	-2.93	2	2	16	116
M	22	D	LP	19	78	38.23	14.49	39.77	14.45	1.54	-3.16	2	2	15	119
E	22	C	LP	07	70	26.68	6.69	29.09	7.13	2.41	-3.33	2	2	21	120
M	32	C	LP	33	75	23.63	8.69	26.13	10.43	2.50	-3.61	1	2	18	123
H	64	C	SO	03	83	43.78	10.62	44.33	12.00	.55	-3.78	1	2	20	125
E	19	A	LP	25	75	28.74	5.76	30.26	6.90	1.52	-4.17	1	2	26	126
E	24	A	SO	07	70	33.50	6.44	34.25	9.26	.75	-4.34	1	2	20	128
M	17	D	LP	07	70	30.06	6.66	30.59	7.35	.53	-4.84	2	2	18	131
E	28	D	LP	27	85	21.18	6.93	22.21	6.32	1.03	-5.09	1	2	17	132
M	16	D	LP	27	85	37.25	14.77	35.50	15.55	-1.75	-6.56	2	2	23	133
E	20	D	LP	33	75	20.35	8.00	19.70	6.63	- .65	-6.88	2	2	21	134
E	16	A	LP	10	73	37.31	14.27	35.31	13.07	-2.00	-6.98	1	1	33	135
E	18	A	LO	05	87	43.17	11.69	39.06	13.07	-4.11	-8.45	2	1	23	136

TABLE 22.--CONTINUED

Grade Level	Pupil N	Schl Code	Schl Dist	MAT Test	High Score	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Raw Gain	Res. Gain	Pretest Group	Gain Curve	Interval Pre-Post	Rank Num.
E	29	A	LP	03	83	42.79	15.90	52.07	15.82	9.28	4.90	3	6	18	1
E	75	A	LP	07	70	30.08	8.36	39.51	11.12	9.43	3.96	3	5	20	3
M	46	A	LP	27	85	43.54	16.59	51.76	14.96	8.22	3.79	3	5	15	4
M	27	A	LP	05	87	35.89	7.89	43.96	11.42	8.07	3.11	3	5	18	6
M	30	A	LP	31	80	37.37	16.43	44.43	18.11	7.07	2.16	3	5	23	7
M	49	D	LP	27	85	40.71	16.14	47.02	16.27	6.31	1.87	3	4	14	8
M	49	A	LP	12	77	24.90	7.93	32.45	9.33	7.55	1.63	3	4	18	10
E	21	C	LP	03	83	40.00	11.76	46.05	13.77	6.05	1.47	3	4	15	12
E	19	A	LP	03	83	46.00	6.93	51.21	8.08	5.21	1.10	3	4	19	14
M	47	A	LP	32	80	55.85	11.73	60.00	9.84	4.15	.61	3	4	18	19
E	30	A	LP	20	73	42.20	8.06	46.97	9.47	4.77	.25	3	4	17	22
E	20	A	LP	03	83	41.55	11.86	46.20	10.89	4.65	.18	3	4	17	23
H	15	D	LP	33	75	50.33	10.79	53.93	10.54	3.60	.01	3	4	15	24
H	24	D	LP	36	60	39.42	4.15	43.75	6.92	4.33	- .22	3	4	15	26
E	23	D	LP	03	83	40.52	13.82	44.65	13.51	4.13	- .38	3	3	16	28
E	53	A	LP	07	70	33.87	8.46	38.36	9.95	4.49	- .65	3	3	19	30
M	47	A	LP	18	90	37.79	9.18	41.91	11.66	4.12	- .67	3	3	18	31
E	24	D	LP	03	83	41.75	6.92	45.17	7.64	3.42	- .92	3	3	19	32
M	46	A	LP	05	87	36.54	9.60	40.46	11.07	3.92	- .98	3	3	18	34
E	33	D	LP	03	83	42.76	11.21	45.67	10.14	2.91	-1.43	3	3	16	39
E	25	A	LP	07	70	37.80	8.82	41.08	9.83	3.28	-1.51	3	3	16	41
M	39	A	LP	09	85	41.15	12.98	44.28	13.32	3.13	-1.57	3	3	21	42
M	30	C	LP	06	75	52.70	13.98	54.50	11.74	1.80	-1.68	3	3	16	43
E	28	D	LP	04	75	29.82	9.15	32.46	8.79	2.64	-2.84	3	3	17	54
M	54	A	LP	11	68	34.43	8.41	36.61	8.82	2.18	-2.91	3	3	18	55
E	19	C	LP	03	83	47.42	13.90	47.58	14.05	.16	-3.73	3	2	18	61
M	54	A	LP	05	87	38.30	9.79	38.43	8.83	.13	-4.62	3	2	20	62
E	25	A	LP	03	83	49.32	11.72	48.40	11.85	- .92	-4.71	3	1	17	64
H	31	A	LP	17	78	49.19	15.08	52.19	14.68	3.00	- .82	4	4	11	02
H	38	A	LP	05	87	66.79	14.20	67.39	15.44	.60	-1.63	4	3	09	05
M	26	D	LO	41	60	30.46	8.09	33.65	8.06	3.19	-2.10	4	2	09	06
H	26	A	LP	08	75	50.42	13.36	36.46	21.47	-13.96	-17.74	4	1	11	07

TABLE 23.--MALE INSTRUMENTAL GAIN SCORES

Grade Level	Pupil N	Schl Code	Schl Dist	MAT Test	High Score	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Raw Gain	Res. Gain	Pretest Group	Gain Curve	Interval Pre-Post	Rank Num.
E	37	A	LO	18	90	39.16	9.42	53.95	9.89	14.79	10.00	2	6	23	3
M	23	D	LO	18	90	47.09	14.83	59.04	15.18	11.95	8.06	1	5	20	5
H	28	C	LO	32	80	32.86	11.17	45.82	12.70	12.96	7.54	2	5	28	6
M	30	C	LO	27	85	48.03	12.64	57.43	10.70	9.40	5.45	1	5	25	10
E	19	A	LO	27	85	47.89	15.26	56.74	14.32	8.85	4.89	2	5	18	13
M	17	D	LP	27	85	30.94	9.87	40.94	10.53	10.00	4.68	2	5	21	14
E	52	A	LO	33	75	36.15	12.44	45.52	10.86	9.37	4.21	1	5	22	19
M	68	A	LO	27	85	55.88	12.03	62.93	10.92	7.05	3.72	1	4	17	25
H	27	A	LO	29	88	42.48	13.60	50.70	14.56	8.22	3.72	1	4	16	26
M	24	A	LO	03	83	44.38	11.97	51.17	10.30	6.79	2.51	1	4	18	33
M	29	D	LP	31	80	42.93	14.19	49.41	13.06	6.48	2.22	2	4	20	36
M	16	C	LP	22	77	36.13	12.80	43.19	16.13	7.06	2.01	2	4	22	38
E	33	A	LO	27	85	49.48	13.63	55.15	11.69	5.67	1.83	2	4	23	40
M	52	A	LO	29	88	50.17	15.60	55.69	13.53	5.52	1.71	1	4	18	41
M	26	A	LO	27	85	62.08	11.85	66.65	9.95	4.57	1.64	2	4	18	42
H	28	A	SO	19	78	55.11	10.53	59.93	14.00	4.82	1.40	1	4	15	45
H	60	A	LO	27	85	70.60	9.13	73.92	6.58	3.32	1.36	2	4	15	46
H	62	A	LO	33	75	63.09	7.21	67.12	3.55	4.03	1.20	2	4	20	48
M	34	C	SO	22	77	43.26	10.71	48.47	9.43	5.21	.90	2	4	21	52
M	37	A	LP	38	72	43.46	12.84	48.70	11.76	5.24	.58	2	4	22	55
M	79	A	SO	33	75	38.75	13.65	44.01	12.69	5.26	.35	2	3	21	60
M	34	A	LO	30	80	51.50	10.68	55.71	8.42	4.21	.31	2	3	22	61
E	71	A	SO	15	72	45.39	10.10	49.87	10.19	4.48	.29	1	3	21	62
M	44	D	SO	07	70	45.00	10.23	48.91	10.89	3.91	- .23	1	3	31	70
M	38	A	LO	27	85	53.97	16.09	58.08	15.01	3.11	- .38	2	3	21	74
H	64	D	SO	32	80	57.30	11.53	59.80	9.88	2.50	- .50	1	3	23	76
H	75	A	SO	27	85	65.20	11.95	67.21	10.59	2.01	- .57	1	3	17	79
H	76	A	LO	33	75	65.16	5.52	67.09	3.67	1.93	- .69	1	3	18	81
M	20	D	LP	22	77	38.40	12.85	42.40	14.47	4.00	- .70	2	3	21	82
H	28	C	SO	33	75	52.89	11.97	55.57	9.33	2.68	- .75	1	3	20	83
H	49	D	SO	39	55	33.47	5.35	37.67	5.47	4.20	- .93	1	3	29	86
H	25	A	LO	27	85	56.68	12.87	58.52	12.43	1.84	-1.21	2	3	20	92
M	21	D	LP	25	75	35.24	11.83	38.81	12.38	3.57	-1.34	1	3	16	98
E	36	A	LO	03	83	58.89	8.69	60.56	9.58	1.67	-1.36	1	3	23	99
H	24	A	SO	23	76	51.25	14.17	53.75	12.74	2.50	-1.40	1	3	21	101
H	39	A	LO	27	85	75.97	6.42	76.08	5.59	.11	-1.51	2	2	24	102

TABLE 23.--CONTINUED

Grade Level	Pupil N	Schl Code	Schl Dist	MAT Test	High Score	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Raw Gain	Res. Gain	Pretest Group	Gain Curve	Interval Pre-Post	Rank Num.
H	68	A	LP	31	80	58.22	11.74	59.65	10.57	1.43	-1.77	1	2	16	103
H	24	A	LO	39	55	31.04	9.59	33.88	9.54	2.84	-2.77	2	2	20	113
M	35	A	LP	14	77	48.43	14.00	49.60	13.60	1.17	-2.89	2	2	21	115
M	83	A	SO	16		45.72	6.28	46.99	6.75	1.27	-3.04	1	2	18	117
M	31	A	LO	42	75	45.71	9.87	46.71	9.56	1.00	-3.14	2	2	15	118
H	15	A	LO	42	75	53.67	10.51	53.60	8.00	- .07	-3.38	1	1	20	122
M	19	A	LP	32	80	59.16	12.27	49.47	12.65	.31	-3.67	2	1	27	124
Mq	23	A	LP	05	87	56.35	13.15	64.35	10.98	8.00	4.79	3	6	15	2
M	44	A	LP	27	85	65.36	14.13	69.48	12.71	4.12	1.53	3	5	20	11
M	23	C	LP	27	85	49.30	11.26	54.22	11.92	4.92	1.13	3	5	17	13
E	33	A	LP	22	77	45.27	10.38	50.42	12.91	5.15	.90	3	5	15	15
M	40	D	LP	03	83	40.88	7.78	46.15	9.47	5.27	.85	3	4	15	16
E	16	A	LO	27	85	47.31	10.88	52.00	11.76	4.69	.73	3	4	18	17
E	19	A	SO	31	80	46.11	13.58	50.68	10.12	4.57	.43	3	4	20	20
M	31	A	LP	31	80	48.81	11.36	53.13	12.43	4.32	.33	3	4	16	21
M	76	A	LP	32	80	53.88	11.28	57.28	10.11	3.40	- .07	3	4	14	25
M	24	A	LP	27	85	36.79	10.61	41.38	15.72	4.59	- .32	3	4	19	27
M	32	A	LP	27	85	57.81	13.50	60.03	12.57	2.22	- .93	3	4	19	33
M	68	A	LP	31	80	55.35	10.60	58.03	10.22	2.68	-1.02	3	4	18	35
M	25	A	SO	37	67	39.12	8.38	42.76	7.79	3.64	-1.08	3	4	14	37
E	20	D	LP	05	87	36.75	11.62	40.20	11.34	3.45	-1.38	3	3	19	38
M	21	D	LP	27	85	32.14	10.09	35.48	12.47	3.34	-1.86	3	3	20	44
E	17	D	LP	10	73	37.59	7.41	40.41	9.76	2.82	-1.94	3	3	18	45
H	57	A	SO	42	75	40.60	9.02	43.11	9.41	2.51	-2.14	3	3	16	47
E	159	C	LP	32	80	37.02	14.78	39.55	15.62	2.53	-2.38	3	3	18	48
H	51	A	LP	46	55	44.90	7.17	46.49	6.13	1.59	-2.73	3	3	16	52
M	23	D	LP	07	70	37.35	7.33	39.35	7.53	2.00	-2.76	3	3	19	53
M	38	C	LP	13	68	42.92	10.85	44.08	12.29	1.16	-3.17	3	2	14	56
M	43	D	LP	07	70	39.72	9.28	41.02	11.80	1.30	-3.22	3	2	14	57
H	41	A	LP	30	80	54.07	9.75	54.02	16.36	- .05	-3.54	3	2	15	58
M	22	C	LP	32	80	59.36	11.32	58.55	10.91	- .81	-3.71	3	2	15	60
M	89	A	LP	33	75	54.72	10.04	57.15	10.19	2.43	-1.04	4	4	16	3

TABLE 24.--MALE CHORAL GAIN SCORES

Grade Level	Pupil N	Schl Code	Schl Dist	MAT Test	High Score	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Raw Gain	Res. Gain	Pretest Group	Gain Curve	Interval Pre-Post	Rank Num.
M	14	D	LP	27	85	23.86	5.69	55.36	10.10	31.50	25.54	2	6	21	1
M	20	A	LP	03	83	48.80	9.51	58.70	8.59	9.90	6.04	1	5	18	9
E	26	A	SO	05	87	57.12	14.68	64.35	13.48	7.23	4.02	2	5	23	20
H	90	A	LO	32	80	41.00	16.38	49.52	15.80	8.52	3.81	2	5	18	23
H	33	C	LP	32	80	54.88	14.97	61.30	13.53	6.42	3.06	1	4	18	30
H	67	D	LO	29	88	46.79	15.59	51.42	16.61	4.63	.64	1	4	25	54
E	24	D	LP	47	50.04		11.07	54.25	10.27	4.21	.55	1	4	18	56
H	35	A	SO	32	80	44.37	16.61	49.06	16.02	4.69	.23	1	4	19	64
H	32	A	LO	30	80	39.78	10.98	44.69	12.53	4.91	.10	1	3	17	66
E	24	C	LP	03	83	50.54	12.29	53.92	14.26	3.38	- .29	2	3	26	72
M	31	A	LO	43	108	42.32	12.45	46.26	13.46	3.94	- .53	2	3	19	78
H	23	A	LO	31	80	44.30	19.06	47.70	17.97	3.40	-1.02	2	3	25	87
H	27	A	SO	27	85	47.89	17.61	50.93	16.16	3.04	-1.27	1	2	26	93
E	23	A	SO	10	73	35.35	8.49	38.22	9.90	2.87	-2.17	2	3	17	108
M	26	A	LO	01	83	46.88	8.17	47.58	9.77	.70	-3.34	2	2	20	121
M	26	C	LO	43	108	67.04	12.78	65.00	12.72	-2.04	-4.27	2	1	16	127
H	14	A	LP	28	90	82.43	5.77	85.14	3.91	2.71	1.67	3	5	16	9
M	49	C	LP	30	80	44.90	13.40	48.51	11.70	3.61	- .61	3	4	18	29
M	53	A	LP	14	77	42.13	11.09	45.70	8.77	3.57	-1.02	3	4	20	36
M	35	A	LP	32	80	41.94	13.21	43.97	14.53	2.03	-2.64	3	3	19	50
H	62	A	LP	28	90	55.29	18.32	56.03	19.99	.74	-2.71	3	3	19	51
E	137	D	SO	01	83	35.67	5.56	36.88	6.31	1.21	-3.65	3	3	10	59
E	57	A	LP	27	85	33.28	11.70	33.86	11.28	.58	-4.62	3	3	21	63
H	56	A	LP	35	68	34.68	12.11	34.88	12.96	.20	-4.97	3	2	13	65
M	22	C	LP	28	90	24.55	10.18	30.00	7.82	5.45	- .49	4	4	12	01
M	23	C	LP	21	82	37.00	13.48	40.52	14.39	3.52	-1.34	4	4	10	04

TABLE 25.--FEMALE INSTRUMENTAL GAIN SCORES

Grade Level	Pupil N	Schl Code	Schl Dist	MAT Test	High Score	Pretest Mean	Pretest SD	Posttest Mean	Posttest SD	Raw Gain	Res. Gain	Pretest Group	Gain Curve	Interval Pre-Post	Rank Num.
M	27	D	SO	33	75	39.00	15.10	47.93	11.95	8.93	4.40	1	6	17	16
M	19	A	LO	07	70	40.63	6.78	47.37	6.80	6.74	2.15	2	5	18	37
H	18	D	SO	31	80	57.94	11.17	60.00	9.27	2.06	- .90	1	4	19	85
M	16	C	LO	42	75	38.69	7.10	42.13	7.20	3.44	-1.28	2	3	18	94
H	71	A	SO	34	63	45.97	6.21	45.63	6.79	- .34	-4.68	1	2	30	129
M	24	A	LP	32	80	54.25	12.32	53.08	14.27	-1.17	-4.78	2	2	20	130
E	35	A	LP	22	77	34.89	8.56	43.51	7.90	8.62	3.55	3	5	21	5
E	30	C	LP	05	87	50.43	11.84	54.80	11.09	4.37	.72	3	4	19	18
M	30	A	LP	28	90	35.53	11.49	39.13	14.17	3.60	-1.43	3	4	18	40
E	20	A	LP	03	83	52.05	8.75	53.50	12.35	1.45	-2.11	3	3	21	46
H	36	A	LP	36	60	41.36	7.00	43.47	7.17	2.11	-2.44	3	3	14	49
H	16	D	LP	24	80	42.88	13.74	41.69	12.52	-1.19	-5.43	3	1	15	66

In order to better interpret the meaning of these data, the music educator sample was stratified into 3 groups -- high, medium, and low -- based on the subjects' G ranked score within pretesting time groups for the 4 teacher type groups: female choral, female instrumental, male choral, and male instrumental. Tables 22 through 25 show the rank order for these groups with two spaces between each success group. While the sample was not trichotomized according to success for the regression analyses which were the major emphases of the study, this tertiary stratification was helpful in describing whether any type of subgroup or grade level group had a tendency to be more successful in G than some other groups.

The first consideration was whether the inclusion of the paid subjects contributed to the total distribution of the sample as far as levels of goodness were concerned. Tables 22 through 25 clearly demonstrate that the paid subjects tended to score in the lower third of the sample in G more than did the unpaid subjects. While many may argue that the paid subjects were really paid volunteers which was true, certain peer and/or administrative pressures were exerted in the school systems in which all the participants participated that the finding that there was less volunteering among the paid subjects was substantiated. Further evidence of this was found in the amount of data sabotage. At least 75 paid subjects or 36 percent of the final sample were coerced into participating in the study by either their music administrators, peer music teachers, or the investigator. Therefore, a volunteer bias as far as the unpaid subjects were concerned appeared evident in that more of these subjects scored higher in G. The selection-specificity effect seemed to be somewhat reduced by obtaining the paid subjects.

Tables 26 and 27 show that grouping by pretesting times was not a factor in music teaching success according to G. The female choral group represented 52 percent of the sample for Group 1-2. Of this figure, 49 percent were in the high group, 50 percent in the middle group, and 58 percent in the low group. This showed a slight weighting factor in the low group but did not deviate much from the total 52 percent figure.

The male instrumental group, while accounting for 32 percent of the sample in Group 1-2, showed only 26 percent in the low group demonstrating that male instrumental music educators tended to be more successful. The same was true for the female choral group in pretesting Group 3. However, when compared to the total sample which was used for the major analyses of the study, grouping by pretesting times did not appear to be a factor in music teaching success as constructed from the pre and posttest administrations of the music achievement tests for the G criterion.

Tables 28 and 29 show the gain score distribution by socioeconomic school categories. The subjects in the female choral category appeared to be more successful in G than those in the disadvantaged category and was the only teacher type group to be so weighted. Since this was the teacher type group with the largest number of subjects, it affected the total sample. The finding is that the residual gain score program did not account for all of the differences in class mean residual gain scores for teachers in socioeconomic categories as had been originally desired. The female choral music educator had an advantage for obtaining success as measured by G if he were teaching in a socially advantaged school. There appeared to be no substantial differences in success for the other music educators according to socioeconomic categories of their schools.

TABLE 26.--GAIN SCORE TOTAL PERCENTAGE DISTRIBUTION BY GROUPS

	FC		FI		MC		MI		Total	
	N	%	N	%	N	%	N	%	N	%
Groups 1-2										
High	22	10.5	2	1.0	5	2.4	16	7.7	45	21.6
Medium	23	11.0	1	.5	7	3.3	15	7.2	46	22.0
Low	26	12.4	3	1.4	4	1.9	12	5.7	45	21.4
Total	71	33.9	6	2.9	16	7.6	43	20.6	136	65.0
Group 3										
High	11	5.2	2	1.0	1	.5	8	3.9	22	10.6
Medium	12	5.7	1	.5	2	1.0	7	3.3	22	10.5
Low	5	2.4	3	1.4	5	2.4	9	4.3	22	10.5
Total	28	13.3	6	2.9	8	3.9	24	11.5	66	31.6
Group 4										
High	1	.5	0		1	.5	0		2	1.0
Medium	1	.5	0		1	.5	1	.5	3	1.4
Low	2	1.0	0		0		0		2	1.0
Total	4	2.0	0		2	1.0	1	.5	7	3.4
Totals										
High	34	16.3	4	1.9	7	3.3	24	11.5	69	33.0
Medium	36	17.2	2	1.0	10	4.8	23	11.0	71	34.0
Low	33	15.8	6	2.9	9	4.3	21	10.0	69	33.0
Total	103	49.3	12	5.8	26	12.4	68	32.5	209	100.0

TABLE 27.--GAIN SCORE WITHIN GROUP PERCENTAGES

	FC		FI		MC		MI		Total	
	N	%	N	%	N	%	N	%	N	%
Group 1-2										
High	22	49.0	2	4.	5	11.	16	35.	45	100.
Medium	23	50.	1	2.	7	15.	15	33.	46	100.
Low	26	58.	3	7.	4	9.	12	26.	45	100.
Total	71	52.	6	4.	16	12.	43	32.	136	100.
Group 3										
High	11	50.	2	9.	1	5.	8	36.	22	100.
Medium	12	55.	1	5.	2	9.	7	31.	22	100.
Low	5	23.	3	14.	5	23.	9	40.	22	100.
Total	28	43.	6	9.	8	12.	24	36.	66	100.
Group 4										
High	1	50.	0		1	50.	0		2	100.
Medium	1	33.3	0		1	33.3	1	33.3	3	99.9
Low	2	100.	0		0		0		2	100.
Total	4	57.	0		2	29.	1	14.	7	100.
Totals										
High	34	49.	4	6.	7	10.	24	35.	69	100.
Medium	36	51.	2	3.	10	14.	23	32.	71	100.
Low	33	48.	6	9.	9	13.	21	30.	69	100.
Total	103	49.	12	6.	26	12.	68	33.	209	100.

Code: FC=Female Choral, FI=Female Instrumental, MC=Male Choral, MI=Male Instrumental

TABLE 28.--GAIN SCORE DISTRIBUTION BY SOCIOECONOMIC CATEGORIES

	FEMALE CHORAL								FEMALE INSTRUMENTAL							
	A		C		D		Total		A		C		D		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
High	27	12.9	4	1.9	3	1.4	34	16.2	2	1.0	1	.5	1	.5	4	2.0
Medium	20	9.5	6	2.9	10	4.8	36	17.2	1	.5	0		1	.5	2	1.0
Low	16	7.7	4	1.9	13	6.2	33	15.8	4	1.9	1	.5	1	.5	6	2.9
Totals	63	30.1	14	6.7	26	12.4	103	49.2	7	3.4	2	1.0	3	1.5	12	5.9

	MALE CHORAL								MALE INSTRUMENTAL							
	A		C		D		Total		A		C		D		Total	
	N	%	N	%	N	%	N	%	N	%	N	%	N	%	N	%
High	4	1.9	2	1.0	1	.5	7	3.4	16	7.7	4	1.9	4	1.9	24	11.5
Medium	5	2.4	3	1.4	2	1.0	10	4.8	15	7.2	2	1.0	6	2.9	23	11.1
Low	7	3.3	1	.5	1	.5	9	4.3	14	6.7	3	1.4	4	1.9	21	10.0
Total	16	7.6	6	2.9	4	2.0	26	12.3	45	21.6	9	4.3	14	6.7	68	32.6

	TOTAL GROUPS							
	A		C		D		Total	
	N	%	N	%	N	%	N	%
High	49	23.4	11	5.3	9	4.3	69	33.0
Medium	41	19.6	11	5.3	19	9.1	71	34.0
Low	41	19.6	9	4.3	19	9.1	69	33.0
Total	131	62.6	31	14.9	47	22.5	209	100.0

BEST COPY AVAILABLE

TABLE 29.--SOCIOECONOMIC GAIN SCORE DISTRIBUTIONS WITHIN GROUP PERCENTAGES

	ADVANTAGED		COMBINATION		DISADVANTAGED		TOTAL	
	N	%	N	%	N	%	N	%
FEMALE CHORAL								
High	27	79	4	12	3	9	34	100
Medium	20	56	6	16	10	28	36	100
Low	16	49	4	12	13	39	33	100
Total	63	61	14	14	26	25	103	100
FEMALE INSTRUMENTAL								
High	2	50	1	25	1	25	4	100
Medium	1	50	0		1	50	2	100
Low	4	67	1	16.5	1	16.5	6	100
Total	7	58	2	17	3	25	12	100
MALE CHORAL								
High	4	57	2	28	1	15	7	100
Medium	5	50	3	30	2	20	10	100
Low	7	78	1	11	1	11	9	100
Total	16	62	6	23	4	15	26	100
MALE INSTRUMENTAL								
High	16	67	4	16	4	16	24	100
Medium	15	65	2	9	6	26	23	100
Low	14	67	3	14	4	19	21	100
Total	45	66	9	13	14	21	68	100

One major concern of the study was whether or not teachers teaching at the elementary level would have a better chance for obtaining G than those music educators teaching at the middle or high school level. For this reason, teachers were instructed to use a group of pupils for the G criterion who were less musically sophisticated than their pupils in the most outstanding performing group. Even though the residual gain statistical procedure accounted for some regression towards the mean effects, it would not be considered to wholly satisfy these requirements. No single statistical procedure is without some error variance.

Table 30 shows the gain score distribution by grade level. The percentages were accumulated across rows for each category: elementary, middle and high school. For example, the female choral group accounted for 33 percent of the high success group; yet, 35 percent of the subjects in this group were elementary and middle school music teachers and only 18 percent were teaching choral music at the high school level. The medium success group showed a definite weighting bias in favor of the high school female choral teachers.

The male instrumental group, too, had a definite weighting advantage in success for the elementary teacher while the reverse was true for the high school teacher who tended to be ranked low in success G. This was true for the female instrumental music educator, too, although there were too few subjects to make adequate interpretations of these findings.

Was grade level per se a contributing factor in determining the music teacher's success in ranked gain score or was the statistic used to derive the class mean residual gain score appropriate enough to account for these grade level differences among success groups? Table 30 appears to portray a definite regression effect: music teachers teaching at the elementary school level where pupils were less musically sophisticated tended to be classified as more successful teachers when the G criterion was interpreted by dividing the sample into thirds: high, medium, and low.

One factor in the investigation was the basic assumption that the music educator was teaching at the grade level and in the speciality most conducive to his achieving success. Grade level per se was not under investigation nor was the music teaching speciality. Was the G criterion really fair to middle and high school music educators? If one standardized music achievement test had been used, the answer to the above question would be negative. There was one additional factor (the most fundamental and important factor) that contributed to the music educator's ability to achieve success on the G criterion. This factor was that the music educator selected the music achievement test used with his pupils. If this test was too easy or too difficult for his children, did not measure adequately the teacher's objectives, or was generally inappropriate contributed to the music educator's ability to produce success that was measurable. In other words, the above findings show that the elementary music teachers tended to pick more appropriate music achievement subtests for measuring their objectives than did their colleagues teaching at the middle and high school levels. It is also possible that the elementary music teachers taught for more content than did their middle and high school colleagues.

One could also find that the music achievement tests did not adequately discriminate among older pupils. This finding would be in contrast to practically all the evidence in existence relating to the Music Achievement Tests, I-IV as reported by Colwell (1967,1970) and Dvorak (1973 in progress). The MAT Tests do discriminate among middle school, high school, and college students. Dvorak has shown that they discriminate among music majors. It appears that the finding is that the elementary music educators selected more appropriate tests and were more successful at achieving their objectives than the middle and high school music educators.

TABLE 30.--GAIN SCORE DISTRIBUTION BY GRADE LEVEL

	HIGH		MEDIUM		LOW		PERCENT	
	N	%	N	%	N	%	N	%
FEMALE CHORAL								
Elementary School	24	35	25	36	20	29	69	100
Middle School	8	35	5	22	10	43	23	100
High School	2	18	6	55	3	27	11	100
Total	34	33	36	35	33	33	103	100
MALE INSTRUMENTAL								
Elementary School	7	58	2	17	3	25	12	100
Middle School	14	38	14	38	9	24	37	100
High School	3	16	7	37	9	47	19	100
Total	24	35	23	34	21	31	68	100
MALE CHORAL								
Elementary School	1	17	2	33	3	50	6	100
Middle School	2	20	5	50	3	30	10	100
High School	3	30	4	40	3	30	10	100
Total	6	23	11	42	9	35	26	100
FEMALE INSTRUMENTAL								
Elementary School	2	67	0		1	33	3	100
Middle School	2	40	1	20	2	40	5	100
High School	0		1	25	3	75	4	100
Total	4	33	2	17	6	50	12	100
TOTAL SAMPLE								
Elementary School	34	38	29	32	27	30	90	100
Middle School	26	35	25	33	24	32	75	100
High School	8	18	18	41	18	41	44	100
Total	68	32.5	72	34.5	69	33	209	100

Table 31 shows the music achievement test distribution for the music teacher type groups within grade level categories. With the exception of tests 40,43,46, and 47, the tests are listed in order of difficulty. High school music educators were advised to select more difficult tests. Slightly more than 50 percent of the high school teachers in the female choral group did not select the more difficult tests; and, therefore, this could have affected their chances for obtaining gain success.

The middle school teachers in the four teacher type groups appeared to be more diverse in selecting music achievement tests. The male choral and male instrumental teachers teaching at the high school level tended to select more difficult tests especially those dealing with music reading: 09,13,14,15,18,21, 22,23,24,25,27,28,29,30,31,32,33,34,35,36,37,38,42,43. Tests 27 through 33 included two music reading subtests (H according to pitch and I according to rhythm). A high school or middle school pupil in band, orchestra, or choir (an elective organization) would be expected to read music. Nevertheless, a music reading objective for this student might have been inappropriate.

It might appear that subtests from the Music Achievement Tests may have been inappropriate for the middle and high school music teachers who primarily directed performing groups (bands, orchestras, and choirs) and, therefore, invalidating the pre and posttesting procedures used to derive the class mean residual gain scores. Some of these music educators may have been forced into believing that these subtests were appropriate and relevant for their music classes either by the school district's music administrator or by the investigator. Considering Colwell's painstaking efforts to validate his tests and determine their reliabilities for use by secondary school music teachers, these latter arguments appear impotent. As previously reported, the Music Achievement Tests do discriminate among secondary school pupils and even among music majors.

TABLE 31.--MUSIC ACHIEVEMENT TEST DISTRIBUTION

	Female C.			Female I.			Male C.			Male I.			
TEST	E	M	H	E	M	H	E	M	H	E	M	H	TOTALS
01	9		1				1	1					12
02	1												1
03	16	1	1	1						1	2		22
04	1						1	1					3
05	5	3	1	1			1			1	1		13
06		1											1
07	10	2	1		1						3		17
08			1										1
09		1											1
10	1						1			1			3
11		2											2
12		1											1
13	1										1		2
14								1			1		2
15	1									1			2
16											1		1
17			1										1
18		1								1	1		3
19		2										1	3
20	1												1
21								1					1
22				1						1	3		5
23												1	1
24						1							1
25	1										1		2
26													0
27	4	4					1	1	1	3	10	4	28
28	2				1			1	2				6
29		1	1						1		1	1	5
30								1	1		1	1	4
31		1				1			1	1	3	1	8
32	1	1	2		1			1	3	1	4	2	16
33	3	1	1		1					1	1	3	11
34						1							1
35									1				1
36			1			1							2
37											1		1
38											1		1
39												2	2
40													0
41		1			1								2
42											1	2	3
43								2					2
44													0
45													0
46												1	1
47	12						1						13
TLS.	69	23	11	3	5	4	6	10	10	12	37	19	209

Code: TLS.=Totals, C.=Choral, I.=Instrumental

At the meetings during which the music teachers were briefed on the study and selected their music testing objectives, the investigator strongly encouraged secondary school music teachers of performing groups to select the more difficult test sections which were pointed out to them if these met their music teaching objectives. The investigator does not feel that these teachers selected the best that they could from the available list of behavioral objectives.

Only one high school band director, who was also the district's music administrator, offered comments to the effect that he thought the tests might not discriminate among his students (and he selected a relatively difficult combination of subtests). Also, one additional school district's music administrator expressed verbally to the investigator that he felt the tests may be inappropriate for all his teachers -- elementary through secondary.

After both pre and posttesting sessions, the music educators were requested to write comments about the Music Achievement Tests. Numerous comments were received and the elementary teachers were the most vocal in highlighting their opinions that the tests were too difficult or possibly inappropriate. A considerable number of positive comments were received from the secondary teachers in contrast to very few negative ones. While it is possible that the secondary school music teachers selected irrelevant tests (perhaps tests that they felt their pupils could answer rather than tests which specifically measured the teacher's objectives), the data from the study neither confirm nor deny the possibility. The data do confirm, however, that selection of appropriate music education objectives and their corresponding evaluative tests may have been more difficult for the secondary school music educators. This being so, it reflects on the music educators' abilities to produce gain and/or select appropriate objectives and tests. Certainly, more research is needed in this area if objective measurements are going to be used as teacher accountability criteria.

The class mean residual gain score criterion G as derived from the pre and posttest music achievement test scores of pupils appeared to be a valid criterion for assessing music teaching success as far as this study was concerned: the gain construct discriminated among the music educators comprising the sample of the study. As with any criterion, certain error variances will be apparent; and as discussed above, these errors may contribute to the ultimate findings of the study.

RATINGS OF MUSIC TEACHING SUCCESS

Five ratings of music teaching success were investigated and comprised the weighted R criterion: pupil, building principal, music supervisor, peer, and self. The Illinois Teacher Evaluation Questionnaire (ITEQ) was used in the study because it was the only published instrument available that was specifically designed for secondary school pupil evaluations of teachers. Permission was granted from the publisher to amend the 40 item ITEQ form to a 23 item form to facilitate the gathering of ratings in the four remaining sources of ratings: principal, supervisor, peer, and self. This was necessary because 17 items related directly to what pupils' thought about the teacher or the class such as the kinds of tests used in the class, the amount of homework assignments, and so forth.

All of the ratings were machine scored. Before any analyses could be accomplished, the ratings were standardized from the original 4 point scale to a 10 point scale with a mean of 5.50 and a standard deviation of 2.00. A data manipulation Standard Scores computer program was used which provided for each music educator one card per rating per teacher for a total of 5 rating evaluation punchcards.

Table 32 shows the homogeneity of the sample on the original four point scale and the need for standardizing the scores. The mean scores were all above 3.10 indicating a tendency for persons to rate the music educators at the successful end of the scale.

TABLE 32.--MUSIC EDUCATOR RATING DATA

Rating Category	N	Mean	SD	Variance	Skewness	Kurtosis
Pupil	255	3.15	.31	.10	-.79	.61
Principal	280	3.33	.49	.24	-.71	.05
Supervisor	226	3.22	.58	.34	-.69	.09
Peer	264	3.48	.33	.11	-.01	1.31
Self	270	3.31	.34	.11	-.59	-.40

While 40 subjects were dropped from the study for incomplete music achievement data, these subjects' rating data could be used to predict missing ratings for those subjects with complete music achievement test data and 3 or more ratings. Rather than use some sort of mean rating score to substitute for missing data, multiple regression equations were formulated which predicted 1 or 2 missing ratings for subjects with missing rating data.

Missing Rating Data Predictions

Ten multiple regression equations were formulated to predict missing ratings in the following categories:

Rating	Number of Subjects Predicted
Pupil	3
Building Principal	8
Music Supervisor	42
Peer Teacher	2
Pupil and Supervisor	1
Pupil and Peer	4
Pupil and Self	1
Principal and Supervisor	2
Supervisor and Peer	1
Peer and Self	2
Total	66

A total of 183 subjects had complete rating data of whom 143 were used in the final analyses. The predicted ratings for 66 subjects brought the total number of subjects for the music educator sample to 209 participants. Table 33 shows the correlation matrix for the rating data.

TABLE 33.--CORRELATION MATRIX FOR MUSIC EDUCATOR RATINGS

N=183	Pupil	Building Principal	Music Supervisor	Peer Teacher	Self
Pupil	1.00				
Principal	.35**	1.00			
Supervisor	.39**	.59**	1.00		
Peer	.40**	.54**	.59**	1.00	
Self	.19**	.21**	.09	.14*	1.00

p<.01** p<,05*

	Mean Score	Standard Deviation
Pupil	5.47	1.97
Principal	5.51	2.04
Supervisor	5.47	2.07
Peer	5.60	1.95
Self	5.53	1.89

All the intercorrelations with the exception of the self-supervisor correlation were significant beyond the .05 level. The mean scores and standard deviations, when rounded, demonstrate that the standard scores were achieved as desired: mean of 5.5 and a standard deviation of 2.00.

Tables 34 through 43 show the regression coefficients for predicting the missing rating data. It is important to note that all of the F ratios for the 10 prediction equations were significant beyond the .05 level. The multiple regression coefficients were respectable ranging from R=.21 to .67 when rounded. These tables show that the building principal, music supervisor and peer teacher ratings contributed the most to the regression equations predictions as indicated by the Beta and b weights. Both are reported for the sake of clarity along with the standard errors, t-tests, degrees of freedom, and significance levels for two tailed tests.

The relationships among the ratings is discussed in the next section of this chapter. The differences suggest three possible domains which were highlighted directly in the factor analysis which weighted the ratings for the R criterion.

BEST COPY AVAILABLE

TABLE 34.--MULTIPLE REGRESSION PUPIL RATING PREDICTIONS

	β Coefficients	b Coefficients	SE of β	SE of b	t-test
Intercept				.55	3.43**
Principal	.06	.06	.09	.84	.71
Supervisor	.22	.21	.08	.08	2.66**
Peer	.24	.24	.08	.08	2.98**
Self	.13	.13	.07	.07	1.86*

Intercept .02 F = 13.40** DF = 4/178

$R = .48$ $R^2 = .23$ SE of Estimate = 1.75

p<.01**, p<.05*

TABLE 35.--MULTIPLE REGRESSION PRINCIPAL RATING PREDICTIONS

	β Coefficients	b Coefficients	SE of β	SE of b	t-test
Intercept				.49	1.11*
Pupil	.05	.05	.06	.07	.71
Supervisor	.41	.40	.07	.07	6.16**
Peer	.30	.31	.07	.07	4.47**
Self	.12	.13	.06	.06	2.15**

Intercept .55 F = 35.40** DF = 4/178

$R = .67$ $R^2 = .44$ SE of Estimate = 1.54

p<.01**, p<.05**

TABLE 36.--MULTIPLE REGRESSION MUSIC SUPERVISOR RATING PREDICTIONS

	β Coefficients	b Coefficients	SE of β	SE of b	t-test
Intercept				.51	2.50**
Pupil	.17	.18	.06	.07	2.66**
Principal	.43	.43	.07	.07	6.16**
Peer	.21	.22	.07	.07	2.93**
Self	-.07	-.07	.06	.06	-1.11

Intercept 1.27 F = .32** DF = 4/178

$R = .65$ $R^2 = .42$ SE of Estimate = 1.60

p<.01**, p<.05*

TABLE 37.--MULTIPLE REGRESSION PEER TEACHER RATING PREDICTIONS

	β Coefficients	b Coefficients	SE of β	SE of b	t-test
Intercept				.49	3.16**
Pupil	.19	.19	.07	.07	2.98**
Principal	.34	.32	.08	.07	4.47**
Supervisor	.22	.21	.08	.07	2.93**
Self	.01	.01	.06	.06	.22

Intercept 1.55 F = 26.53** DF = 4/178

$R = .61$ $R^2 = .37$ SE of Estimate - 1.56

p<.01**, p<.05*

TABLE 38.--MULTIPLE REGRESSION PUPIL AND SUPERVISOR PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-test	
	Pupil	Supervisor	Pupil	Supervisor	Pupil	Supervisor	Pupil	Supervisor	Pupil	Supervisor
Intercept							.54	.49	4.14**	3.39**
Principal	.16	.46	.16	.46	.08	.07	.08	.07	2.05**	6.53**
Peer	.30	.26	.30	.27	.08	.07	.08	.07	3.77**	3.73**
Self	.11	-.05	.12	-.05	.07	.06	.08	.07	1.68*	- .77
Intercept	2.23	1.67								
F =	14.00**	38.63**								
R =	.45	.63								
SE of Est =	1.78	1.62								
R^2 =	.20	.39								

DF = 3/179

p<.01**, p<.05*

TABLE 39.--MULTIPLE REGRESSION PUPIL AND PEER PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-test	
	Pupil	Peer	Pupil	Peer	Pupil	Peer	Pupil	Peer	Pupil	Peer
Intercept							.53	.47	4.44**	4.23**
Principal	.15	.37	.15	.35	.08	.08	.08	.07	1.79*	4.80**
Supervisor	.29	.28	.28	.27	.08	.07	.08	.07	3.52**	3.73**
Self	.14	.04	.14	.04	.07	.07	.06	.07	1.96	.64
Intercept	2.36	2.01								
F =	14.28**	31.05**								
R =	.44	.59								
R^2 =	.19	.34								
SE of Est =	1.79	1.59								

DF = 3/179

p<.01**, p<.05*

TABLE 40.--MULTIPLE REGRESSION PUPIL AND SELF PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-test	
	Pupil	Self	Pupil	Self	Pupil	Self	Pupil	Self	Pupil	Self
Intercept							.45	.48	5.45**	9.36**
Principal	.09	.22	.09	.20	.09	.10	.08	.09	1.03*	2.28**
Supervisor	.22	-.07	.21	-.07	.08	.09	.08	.09	2.54**	- .77
Peer	.25	.06	.25	.06	.08	.09	.08	.08	3.05**	.64
F =	16.48**	2.98*								
R =	.47	.22								
R^2 =	.22	.05								
SE of EST =	1.76	1.86								
Intercept	2.46	4.46								

p<.01**, p<.05*

TABLE 41.--MULTIPLE REGRESSION PRINCIPAL AND SUPERVISOR PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-test	
	Principal	Supervisor	Principal	Supervisor	Principal	Supervisor	Principal	Supervisor	Principal	Supervisor
Intercept							.53	.55	2.44**	3.32**
Pupil	.14	.23	.14	.24	.07	.07	.07	.07	3.05**	3.31**
Peer	.47	.41	.49	.43	.07	.07	.07	.07	6.92**	5.90**
Self	.12	-.02	.13	-.12	.06	.06	.07	.07	1.85*	- .25

DF = 3/179

	Principal	Supervisor
Intercept	1.23	1.83
F Ratio	28.63**	24.57**
R =	.57	.54
R^2 =	.32	.30
SE of Est.	1.70	1.75

p<.01**, p<.05*

TABLE 42.--MULTIPLE REGRESSION SUPERVISOR AND PEER PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-test	
	Supervisor	Peer	Supervisor	Peer	Supervisor	Peer	Supervisor	Peer	Supervisor	Peer
Intercept							.50	.48	3.40**	3.92**
Pupil	.22	.25	.23	.24	.06	.07	.07	.06	3.52**	3.77**
Principal	.52	.45	.53	.43	.06	.07	.06	.06	8.28**	6.92**
Self	-.07	-.00	-.07	-.00	.06	.06	.07	.06	-1.10	- .02

DF = 3/179

	Supervisor	Peer
Intercept	1.69	1.90
F Ratio	37.84**	31.19**
R =	.62	.59
R^2 =	.39	.34
SE of Est.	1.63	1.59

p<.01**, p<.05**

TABLE 43.--MULTIPLE REGRESSION PEER AND SELF PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-test	
	Peer	Self	Peer	Self	Peer	Self	Peer	Self	Peer	Self
Intercept							.41	.49	3.88**	8.44**
Pupil	.20	.16	.20	.15	.07	.08	.06	.08	3.05**	1.96*
Principal	.34	.21	.32	.20	.07	.09	.07	.08	4.59**	2.35**
Supervisor	.22	-.10	.21	-.92	.08	.09	.07	.08	2.93**	-1.10

DF = 3/179

	Peer	Self
Intercept	1.60	4.13
F Ratio	35.55**	4.19**
R =	.61	.26
R^2 =	.37	.07

	Peer	Self
SE of Est.	1.56	1.85

p<.01**, p<.05*

Music Student Teacher Ratings

Table 44 shows the mean scores for the music student teacher ratings for the 4 point scale before this scale was standardized to a 10 point scale with a mean of 5.50 and a standard deviation of 2.00. As was the case with the music educator sample, the music student teacher ratings were homogeneous with the lowest mean rating of 3.00 when rounded.

TABLE 44.--MUSIC STUDENT TEACHER RATING MEAN SCORES

RATING	N	MEAN	STANDARD DEVIATION	VARIANCE	SKEWNESS	KURTOSIS
Pupil	71	3.04	.28	.08	-.24	-.38
Coop. Teacher	66	3.11	.52	.27	-.78	-.05
Superv. Teacher	76	3.07	.60	.36	-.57	-.27
Peer	86	2.95	.37	.14	-.43	.73
Self	75	3.28	.30	.08	.05	-.43

The ratings for the music student teacher cross validation sample were less significantly interrelated as shown by Table 45. The significance of the correlations among the 5 ratings relates to how well missing data may be predicted using multiple regression techniques. In most cases, the more significant the relationships more accurate predictions will result. The cooperating teacher-pupil correlation was significant beyond the .05 level: supervisor-cooperating teacher, peer-supervising teaching, and self-peer rating beyond the .01 level. The other correlations were not statistically significant for a two tailed test.

The scores were standardized as desired with a slight exception for the cooperating teaching rating mean which was 6.00. Otherwise, the ratings satisfied the requirements of a mean of 5.5 and standard deviation of 2.00. This procedure standardized the ratings of the music student teacher sample in the similar manner in which the ratings for the music educator sample were standardized thus facilitating the use of the music student teacher sample for cross validation purposes.

TABLE 45.--MUSIC STUDENT TEACHER RATING CORRELATION MATRIX

N=35	Pupil	Cooperating Teacher	Supervisor	Peer	Self
Pupil	1.00				
Coop. Teacher	.33*	1.00			
Supervisor	.29	.53**	1.00		
Peer	-.00	.22	.51**	1.00	
Self	.11	.16	.15	.49**	1.00

p<.01**, p<.05*

Music Student Teacher Rating Means and Standard Deviations

	Means	Standard Deviations
Pupil	5.70	1.95
Cooperating Teacher	6.00	1.70
Supervising Teacher	5.89	1.96
Peer	5.80	1.60
Self	5.27	1.94

Table 46 shows that 12 multiple regression equations were formulated for predicting missing rating data for 38 music student teachers from the originally complete rating data sample of 35 subjects. Of interest is that 2 more equations were required for this sample than for the music educator sample. Forty-five percent of the missing ratings were due to the cooperating teacher: 10 cooperating teachers failed to rate their student teachers, 5 missing ratings were due to no pupil and cooperating teacher ratings, and 2 missing ratings were due to no cooperating teacher, supervising teacher, and self ratings. At least two reminders were mailed to the cooperating and supervising teachers and included additional rating blanks. Still, the fact that more ratings were predicted than received raises a serious rating or criterion validity question. The fact that these missing ratings were not gathered reduced the usefulness of the music student teacher cross validation sample. The music student teachers' rating data were questionable as far as validity was concerned.

Because many of the intercorrelations were insignificant statistically, the multiple regression equations which predicted the missing ratings must be viewed cautiously. Where most of the missing rating data occurred, the F ratios were significant: Table 47 for predicting the cooperating teacher ratings beyond the .05 level, Table 49 for predicting the supervising teacher ratings beyond

TABLE 46.--MUSIC STUDENT TEACHER PREDICTIONS FOR MISSING DATA

Predicted Rating 12 REGRESSION EQUATIONS	Number
Pupil	4
Cooperating Teacher	10
Supervising Teacher	8
Peer	2
Pupil and Cooperating Teacher	5
Pupil and Supervising Teacher	1
Pupil and Self	1
Cooperating Teacher and Supervising Teacher	1
Cooperating Teacher and Self	1
Supervising Teacher and Peer	1
Supervising Teacher and Self	2
Peer and Self	2
Total	38

the .01 level, and Table 51 for predicting the pupil and cooperating teacher beyond the .01 level for the cooperating teacher ratings but insignificant for the pupil ratings. Of those tables showing the predictions for 2 missing ratings, only Table 55 (cooperating teacher and self ratings), Table 56 (student teacher supervisor and peer ratings), and Table 57 (supervisor and self ratings) had significant F ratios beyond the .05 level for both predicted missing ratings.

The criterion data for the music student teacher sample were not as valid as originally desired. The finding is that the music student teacher criterion data were substantially biased and open to serious error variance.

Tables47 through 58 show the student teacher multiple regression equations' predictions for missing rating data. These tables further highlight the fact that the equations while satisfying the statistical designs and requirements of the study were not perforce as valid or reliable as the predictions for the music educator sample.

TABLE 47.--MULTIPLE REGRESSION PUPIL STUDENT TEACHER PREDICTIONS

	β Coefficients Pupil	b Coefficients Pupil	SE of β Pupil	SE of b Pupil	t-test Pupil
Intercept				1.54	2.37*
Coop. Tchr.	.21	.23	.20	.23	1.03'
Supervisor	.29	.29	.23	.23	1.29
Peer -	-.28	-.34	.22	.27	-1.26
Self	.17	.17	.19	.19	.88

DF =4/30

	Pupil
Intercept	3.65
F Ratio	1.56
R	.41
R^2	.17
SE of Est.	1.89

p<.01**, p<.05*

TABLE 48.--MULTIPLE REGRESSION COOPERATING TEACHER PREDICTIONS

	β Coefficients Coop. Tchr.	b Coefficients Coop. Tchr.	SE of β Coop. Tchr.	SE of b Coop. Tchr.	t-test Coop. Tchr.
Intercept				1.24	2.12*
Pupil	.17	.15	.16	.14	1.03
Supervisor	.52	.45	.19	.16	2.78**
Peer	-.11	-.11	.20	.22	- .52
Self	.12	.10	.18	.15	.66

DF =4/30

	Coop. Tchr.
Intercept	2.61
F Ratio	3.68*
R	.57
R^2	.33
SE of Est.	1.49

p<.01**, p<.05*

TABLE 49.--MULTIPLE REGRESSION SUPERVISING TEACHER PREDICTIONS

	β Coefficients Supervisor	b Coefficients Supervisor	SE of β Supervisor	SE of b Supervisor	t-tests Supervisor
Intercept				1.32	- .41
Pupil	.18	.18	.14	.14	1.29
Coop. Teacher	.39	.45	.14	.16	2.78**
Peer	.51	.63	.15	.19	3.39**
Self	-.18	-.18	.15	.15	-1.20
	Supervisor				DF=4/30
Intercept	- .54				
F Ratio	7.36**				
R	.70				
R^2	.50				
SE of Est.	1.48	p<.01**,	p<.05*		

TABLE 50.--MULTIPLE REGRESSION PEER STUDENT TEACHER PREDICTIONS

	β Coefficients Peer	b Coefficients Peer	SE of β Peer	SE of b Peer	t-tests Peer
Intercept				1.00	2.63*
Pupil	-.18	-.15	.14	.12	-1.26
Coop. Teacher	-.08	-.78	.16	.15	- .52
Supervisor	.54	.44	.16	.13	3.39**
Self	.44	.36	.14	.11	3.22**
	Peer				DF=4/30
Intercept	2.62				
F Ratio	6.65**				
R	.69				
R^2	.47				
SE of Est.	1.24	p<.01**,	p<.05*		

TABLE 51.--MULTIPLE REGRESSION PUPIL AND COOPERATING TEACHER PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-tests	
	Pupil	Coop. T.	Pupil	Coop. T.	Pupil	Coop. T.	Pupil	Coop. T.	Pupil	Coop. T.
Intercept							1.35	1.07	3.26**	3.05**
Supervisor	.42	.59	.41	.51	.20	.18	.19	.15	2.13*	3.37**
Peer	-.31	-.16	-.38	-.17	.22.	.20	.27	.21	-1.41	- .79
Self	.20	.15	.20	.13	.19	.17	.19	.15	1.05	.87

DF = 3.31

	Pupil	Coop. T.
Intercept	4.41	3.26
F Ratio	1.72	4.55**
R	.38	.55
R^2	.14	.31
SE of Est.	1.89	1.49

p<.01**, p<.05*

TABLE 52.--MULTIPLE REGRESSION PUPIL AND SUPERVISING TEACHER PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-tests	
	Pupil	Supervisor	Pupil	Supervisor	Pupil	Supervisor	Pupil	Supervisor	Pupil	Supervisor
Intercept							1.56	1.23	2.36*	.10
Coop. T.	.34	.45	.39	.52	.17	.13	.19	.15	1.96	3.37**
Peer	-.14	.49	-.17	.60	.19	.15	.24	.19	- .70	3.22**
Self	.12	-.16	.12	-.6	.19	.15	.19	.15	.65	-1.05

DF = 3/31

	Pupil	Supervisor
Intercept	3.69	.12
F Ratio	1.49	9.17**
R	.35	.68
R^2	.13	.47
SE of Est.	1.91	1.50

p<.01**, p<.05*

TABLE 53.--MULTIPLE REGRESSION PUPIL AND SELF STUDENT TEACHER PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-tests	
	Pupil	Self	Pupil	Self	Pupil	Self	Pupil	Self	Pupil	Self
Intercept							1.51	1.40	2.59*	1.07
Coop. T.	.23	.16	.27	.18	.20	.18	.22	.21	1.18	.89
Supervisor	.26	-.22	.25	-.22	.22	.21	.22	.21	1.15	-1.05
Peer	-.18	.56	-.22	.68	.19	.18	.24	.22	- .95	3.13**

DF = 3/31

	Pupil	Self
Intercept	3.91	1.50
F Ratio	.183	3.73*
R	.39	.52
R^2	.15	.27
SE of Est.	1.88	1.74

p<.01**, p<.05*

TABLE 54.--MULTIPLE REGRESSION COOPERATING TEACHERS AND SUPERVISOR PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-tests	
	Coop. T.	Supervisor	Coop. T.	Supervisor	Coop. T.	Supervisor	Coop. T.	Supervisor	Coop. T.	Supervisor
Intercept							1.36	1.36	2.20*	.59
Pupil	.33	.31	.29	.31	.17	.14	.15	.15	1.97	2.13*
Peer	.20	.60	.22	.73	.19	.16	.20	.20	1.08	3.62**
Self	.03	-.17	.02	-.17	.19	.17	.17	1.7	.15	-1.03

DF = 3/31

	Coop. T.	Supervisor
Intercept	2.98	.80
F Ratio	1.92	5.95**
R	.40	.60
R^2	.16	.37
SE of Est.	1.64	1.64

p<.01**, p<.05*

BEST COPY AVAILABLE

TABLE 55.--MULTIPLE REGRESSION COOPERATING TEACHER AND SELF PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-tests	
	Coop. T.	Self	Coop. T.	Self	Coop. T.	Self	Coop. T.	Self	Coop. T.	Self
Intercept							1.21	1.42	2.28*	.92
Pupil	.19	.17	.16	.17	.16	.16	.14	.16	1.18	1.05
Supervisor	.50	-.19	.43	-.19	.18	.19	.16	.19	2.73**	-1.03
Peer	-.04	.58	-.40	.71	.17	.18	.17	.22	- .21	3.23**

	Coop. T.	Self
Intercept	2.75	1.31
F Ratio	4.85**	3.89*
R	.57	.52
R^2	.32	.27
SE of Est.	1.47	1.74

DF = 3/31

p<.01**, p<.05*

TABLE 56.--MULTIPLE REGRESSION STUDENT TEACHER SUPERVISOR AND PEER PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-tests	
	Supervisor	Peer	Supervisor	Peer	Supervisor	Peer	Supervisor	Peer	Supervisor	Peer
Intercept							1.35	1.13	1.35	2.92**
Pupil	.12	-.11	.12	-.93	.16	.16	.16	.13	.76	- .70
Coop. T.	.48	.18	.56	.17	.17	.16	.18	.15	3.02**	1.08
Self	.06	.47	.06	.39	.15	.16	.15	.13	.40	3.00**

	Supervisor	Peer
Intercept	1.53	3.30
F Ratio	4.47**	3.75*
R	.55	.52
R^2	.30	.27
SE of Est.	1.72	1.43

DF = 3/31

p<.01**, p<.05*

TABLE 57.--MULTIPLE REGRESSION SUPERVISOR AND SELF STUDENT TEACHER PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-tests	
	Supervisor	Self	Supervisor	Self	Supervisor	Self	Supervisor	Self	Supervisor	Self
Intercept							1.32	1.56	- .57	.71
Pupil	.16	.11	.16	.11	.14	.17	.14	.17	1.15	.65
Coop. T.	.39	.02	.45	.03	.14	.17	.16	.19	2.73**	.15
Peer	.43	.48	.52	.58	.13	.16	.16	.19	3.18**	3.00**

DF = 3/31

	Supervisor	Self
Intercept	- .74	1.11
F Ratio	9.20**	3.43**
R	.69	.50
R^2	.47	.25
SE of Est.	1.49	1.76

$p<.01$**, $p<.05$*

TABLE 58.--MULTIPLE REGRESSION PEER AND SELF STUDENT TEACHER PREDICTIONS

	β Coefficients		b Coefficients		SE of β		SE of b		t-tests	
	Peer	Self	Peer	Self	Peer	Self	Peer	Self	Peer	Self
Intercept							1.03	1.46	3.85**	2.59*
Pupil	-.15	.06	-.13	.06	.16	.19	.13	.19	- .95	.30
Coop. T.	-.14	.10	-.04	.12	.18	.21	.17	.24	- .21	.47
Supervisor	.58	.08	.47	.08	.18	.21	.15	.21	3.18	.40

DF = 3/31

	Peer	Self
Intercept	3.98	3.77
F Ratio	4.14*	.39
R	.53	.19
R^2	.29	.04
SE of Est.	1.41	1.00

$p<.01$**, $p<.05$*

Music Educator Rating Analysis

Before the R criterion could be generated, it was necessary to investigate possible differences in rating evaluations that could be attributed to music teacher type (instrumental or choral) and grade level (elementary, middle, and high school). Analysis of variance was used to test the null hypothesis that there would be no significant differences in rating evaluations among these teacher type grade level groups. Teachers were stratified into 6 groups according to teacher type and grade level: (1) elementary instrumental, (2) elementary choral, (3) middle school instrumental, (4) middle school choral, (5) high school instrumental, and (6) high school choral. A 6 by 5 analysis of variance was conducted.

The following table shows that there were no significant differences beyond the .05 level among the 6 groups of music educators for the 5 ratings of music teaching success under investigation: pupil, building principal, music supervisor, peer teacher, and self. The generation of the factorially derived R criterion could proceed. There were no advantages in rating evaluations attributed to teachers in any of the 6 categories. Furthermore, pupil ratings did not differ for the 3 levels of public schools and the teacher type groups. The null hypothesis that there were no significant differences among the 6 teaching groups failed to be rejected because no differences were found beyond the .05 level for a two tailed test.

TABLE 59.--ANALYSIS OF VARIANCE FOR RATINGS

Source	Nesting	Denominator	Degrees of Freedom		Sum of Squares	SS Digits	Mean Squares	F Ratio	Probability
			Num.	Den.					
B		A	5	202	43.81	*8*	8.76	.95	.45
A	B		202		1863.42	*8*	9.22		
C		AXC	4	808	5.32	*8*	1.33	.59	.67
BXC		AXC	20	808	44.53	*7*	2.23	.98	.48
AXC	B		808		1832.19	*7*	2.27		

A = Subjects
B = the 6 teacher type-grade level groups
C = the 5 ratings of music teaching success: pupil, principal, supervisor, peer, and self

Factor A is the replication factor and A has an unequal number of levels for each combination of levels of the factors in which it is nested. The number of levels is proportional and hence the design is balanced. The Analysis of Variance is exact except for truncation and rounding errors.

Generating the R Criterion

The weights from the first principal component of a factor analysis were used to generate factor scores which, in turn, generated a composite rating R criterion score per subject. The complete rating data after predicting missing ratings were submitted to factor analysis without rotation. From the first principal component, a factor scores program was used (second order factor analysis) which weighted each of the 5 ratings and generated a single R criterion score. Because the missing data rating predictions were included so that all participants had complete rating data, the resulting means, standard deviations, and correlations as shown below were slightly different from those used to generate the missing data scores.

TABLE 60.--MUSIC EDUCATOR RATING CORRELATION MATRIX

N=213	Pupil	Principal	Supervisor	Peer Teacher	Self
Pupil	1.00				
Principal	.27**	1.00			
Supervisor	.34**	.63**	1.00		
Peer	.34**	.55**	.57**	1.00	
Self	.26**	.19**	.14	.20	1.00

	Mean	Standard Deviation	
Pupil	5.24	2.26	
Principal	5.64	1.94	
Supervisor	5.56	1.85	
Peer	5.61	1.90	
Self	5.46	1.88	$p<.01$**, $p<.05$*

The correlations, while varying somewhat from the complete matrix as presented before, were all within the same degree and direction. The same was true for the means and standard deviations which maintained the desired mean of 5.50 and the standard deviation of 2.00. Pupil rating mean score was the lowest while the standard deviation was the highest.

The "Principal Components-Principal Axis Factor Analysis Program" was used to generate a rating factor score for each participating subject, the R criterion. Because there were 5 variables or ratings to be factor analyzed, 5 factors were requested of which the first factor from the factor analysis was used to generate the factor scores. The following table presents the factor analysis data for the music educator sample.

TABLE 61.--FACTOR ANALYSIS OF 5 RATINGS

Factor	Eigenvalues	Percent Variance	Cumulative Percent
1	2.48	49.61	49.61
2	.99	19.82	69.42
3	.71	14.23	83.65
4	.46	9.24	92.89
5	.36	7.11	100.00

Trace = 5.00
The sume of the first 5 roots = 5.00

COLUMN OF EIGENVECTORS

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Pupil	.38	.42	-.81	.12	.13
Principal	.51	-.26	.27	.41	.65
Supervisor	.52	-.28	.03	.32	-.74
Peer	.51	-.15	.14	-.84	.05
Self	.25	.81	.52	.02	-.10

FACTOR MATRIX

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Pupil	.59	.41	-.68	.08	.08
Principal	.81	-.26	.22	.28	.39
Supervisor	.83	-.28	.03	.22	-.44
Peer	.80	-.15	.14	-.57	.03
Self	.39	.81	.44	.02	.06

Table 61 helps clarify the rating relationships. Because the primary purpose of the factor analysis was to generate factor scores for the R criterion, it was unnecessary to use a varimax rotation to assist in locating possible subdomains within the rating criteria although this technique would have been helpful if the emphasis of the study had been an investigation of the criterion domain. The unrotated factor matrix substantiated some probable dimensions of the ratings within the 5 rating categories.

Factor 1 was probably an adult rating factor as indicated by the closeness of the weights given to the supervisor, principal and peer rating variables. The music supervisor was weighted slightly higher than the building principal and peer ratings; yet, these 3 ratings were very close and clustered together.

Factor 2 appeared to be a self rating factor while Factor 3 was probably the pupil rating dimension or domain. Factors 4 and 5 were less distinct but appeared to be peer and supervisor dimensions respectively.

While one might originally find that the ratings were rather equally balanced as shown by their highly significant intercorrelations, the finding from the factor analysis was that within these 5 rating categories, three subdomains appeared to exist and were weighted appropriately for the R criterion. The order of weighting was: music supervisor, building principal, and peer teacher (the adult rating factor), the pupil rating factor which received considerably less weight, and the self rating factor or domain which received the least amount of weight in deriving the R criterion.

That the music supervisor, building principal, and peer teacher ratings would be so closely weighted was an unexpected finding. Intuitively, one would feel that the music supervisor would be able to best judge the effectiveness of a music educator and that a building principal and peer teacher might have more difficulty evaluating a music teacher because of varying frames of reference from which to judge. Furthermore, the peer teacher may have never seen the music educator teach music while it was presumed that the music supervisor and building principal had observed the music educator in the classroom. This finding suggested the possibility that peer teacher ratings may be as valid and useful as music supervisor and building principal ratings. Furthermore, it appeared that the music supervisors and building principals rated the music educators honestly. One might expect that these administrators had a stake in the music educator's success since they probably had something to do with the music educator being hired in the first place. Nevertheless, this factor did not appear to show in the data. This could also suggest that these three ratings were invalid and based on a third factor such as personality, dependability, or some other kind of attitudinal variable. Comments from administrators tended to support this finding.

Another interesting sidelight to the findings was that the peer teacher evaluations and the self ratings were quite dissimilar. This suggested that the music educators were honest in obtaining the cooperation of their colleagues in rating them and that the music educators did not complete the peer evaluations themselves which was a possibility in the study.

From the above data analysis, it was apparent that the pupils were rating the music teachers from a different frame of reference. This could be considered to be a pupil process type of attitudinal evaluative variable criterion. In any case, the factor analysis suggested the possibility that pupil rating was a separate domain of evaluation and, perhaps, should be treated as such.

The self rating domain indicated a considerably different weighting factor when compared to the other ratings. This was the music educator's perceived view of his own success. It appeared from the evidence that the music educator did not see himself in the same evaluative light as did his pupils or his fellow adult colleagues.

The finding from the above data was that there were 3 probable dimensions operating within the rating criteria. These were weighted appropriately according to the first principal component through the use of a least squares factor scores procedure, a type of secondary factor analysis. Had the 5 ratings been assumed to be of equal value and merit, a completely different rank order of R would have occurred. The investigator decided to weight the ratings by using the best technique available which was the factor analytic technique. Once the R score was generated for each subject, the sample was rank ordered from high to low according to the R criterion score.

Music Student Teachers

The same approach was used with the music student teacher cross validation sample. Below is the correlation matrix for the sample. These figures include the missing data predicted scores and are somewhat inflated from those reported previously.

TABLE 62.--MUSIC STUDENT TEACHER CORRELATION MATRIX

N=76	Pupil	Cooperating Teacher	Supervisor	Peer	Self
Pupil	1.00				
Coop. Teacher	.34**	1.00			
Supervising T.	.29**	.61**	1.00		
Peer	-.04	.30**	.59**	1.00	
Self	-.02	.09	-.00	.16	1.00

p<.01**, p<.05*

	Mean	Standard Deviation
Pupil	4.65	2.76
Cooperating Teacher	5.59	1.89
Supervising Teacher	5.51	2.18
Peer	5.73	2.07
Self	5.40	1.96

By including the missing data rating prediction scores, one more significant correlation was found - peer to cooperating teacher. With the exception of the self-pupil rating correlation, the correlations were all in the same direction. The differences of Table 62 from Table 45 were that both cooperating teacher and supervising teacher ratings were significantly related to pupil ratings beyond the .01 level and that the self-peer rating correlation was not significant when it originally was significant. It is interesting to note that the supervisor and cooperating teacher ratings were significantly related beyond the .01 level for a two tailed test. The same was true for the peer-college supervisor correlation.

The means and standard deviations remained relatively stable with the exception of the pupil rating which changed from a mean of 5.70 and a standard deviation of 1.95 to a mean of 4.65 and a standard deviation of 2.76. Otherwise, the design requirements of the study were maintained and somewhat close to the desired means of 5.50 and standard deviations of 2.00.

Table 63 shows the music student teacher factor analysis data. While the figures were somewhat different than the music educator sample, they were all in the same general direction.

TABLE 63.--MUSIC STUDENT TEACHER FACTOR ANALYSIS DATA

Factor	Eigenvalues	Percent Variance	Cumulative Percent
1	2.14	42.85	42.85
2	1.14	22.82	65.67
3	.94	18.71	84.38
4	.52	10.49	94.87
5	.26	5.13	100.00

COLUMNS OF EIGENVECTORS

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Pupil	-.30	.64	.44	.53	.15
Cooperating Teacher	-.55	.17	.12	- .72	.37
Supervisor	-.61	-.00	-.22	.04	-.76
Peer	-.47	-.49	-.32	.44	.49
Self	-.10	-.57	.80	.01	-.15

FACTOR MATRIX

Variable	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Pupil	-.45	.68	.42	.39	.08
Cooperating Teacher	-.80	.18	.12	-.52	.19
Supervisor	-.90	-.00	-.21	.03	-.38
Peer	-.68	-.52	-.31	.32	.25
Self	-.15	-.61	.77	.01	-.18

The weighting of the 5 ratings while similar to the music educator sample was somewhat different in degree. The above data support the finding that there were 4 possible subdomains within the 5 ratings. The first domain suggested an adult rating dimension as reflected by the closeness in weightings for the supervising and cooperating teacher ratings. Like the music educator sample, the college music supervisor rating received the most weight followed closely by the cooperating teacher. This suggested that both college supervisors and cooperating teachers rate music student teachers from the same frame of reference.

The peer student teacher differed from the music educator peer rating in that much less weight was given this rating. The peer student teachers did not rate from the same frames of reference as the supervising and cooperating teachers. Therefore, these data support the finding of a separate peer or student colleague domain even though this did not appear as a separate factor per se in the factor matrix. Better defined were (1) the cooperating-supervising teacher dimension which appeared to be Factor 1, (2) the pupil dimension which appeared in Factor 2, (3) the self rating dimension which was less well defined in Factor 3, (4) a cooperating teacher dimension which was found in Factor 4, and (5) a supervising teacher dimension was located in Factor 5.

The weightings for the music student teacher cross validation sample were as follows: supervising teacher and cooperating teacher ratings followed by the peer student teacher rating with the pupil ratings fourth and the self ratings last. The order was the same as for the music educator sample with the peer rating factor becoming a separate domain and receiving considerably less weight in the music student teacher sample. The above factor analytic procedures generated factor scores for each participating music student teacher which were his R criterion score. The sample was rank ordered from high to low in R.

Summary

Factor analytic techniques were used to weight the ratings of music teaching and music student teaching success. Both samples were rank ordered according to the factor rating score which became the R criterion score. While the music educator sample data showed 3 rating evaluation domains (adult, pupil, and self), the music student teacher sample had 4 possible but less well defined domains (adult, peer student teacher, pupil and self). There was also the possibility of a peer-self domain.

The above procedures generated 4 criterion variables as constructs of music teaching success for the music educator sample: G (gain), R (rating), G+R (gain plus rating), and G-R (gain minus rating), a discrepancy score. The music educator sample was rank ordered within the G and R categories. G+R added the R ranked score to the G ranked score for a composite criterion score while G-R subtracted the R ranked score from the G ranked score for a discrepancy criterion score.

RELATIONSHIPS AMONG ACCOUNTABILITY CRITERIA

The first subproblem of the study referred to the relationships among teacher accountability criteria used to define music teaching success:

a. class mean residual gain scores of pupils
b. teacher ratings by pupils
c. teacher ratings by principals
d. teacher ratings by supervisors
e. teacher ratings by peers
f. teacher self ratings

The previous discussion has highlighted relationships within the gain score data as far as pretesting time differences were concerned and the interrelationships among the rating data. This section presents the findings for the relationships among the gain and rating criterion data.

Spearman's Rho

The first analysis was undertaken in order to investigate the relationship between the class mean residual gain score rank G and the factor analytically derived rating score rank R. Spearmen's rho, a rank correlation coefficient which is a variant of the Pearson Product Moment Correlation Coefficient, was used to initially investigate the null hypothesis that there were no significant relationships between G and R beyond the .05 level of significance. While the regular Pearson formula was used in later analyses, the investigator felt that it was important to first check the relationship for the rank ordered data within music educator teacher type subgroups. Dubois emphasizes a caution, however, for using rho:

> It is to be noted that the distribution of ranks, like the distribution of percentiles, is rectilinear. Accordingly, if a continuous variable is ranked prior to finding the correlation, some information is lost. In general, p will be a little smaller than r for the same data, but the difference is trifling. Since a major function of p is to obtain an estimate of the correlation from a small sample of cases, the difference between p and r is inconsequential (Dubois 1965:229).

The formula for rho which was used for each of the 4 subsamples follows:

$$p = 1 - \frac{6\Sigma D^2}{N(N^2-1)}$$

In the above formula, N is the number of ranked cases and ΣD^2 is the sum of the squares of the differences in ranks. The following are the findings for the correlations between G and R as estimated by rho for the following teacher type groups, none of which were significant beyond the .05 level for a two tailed test:

Group	Number	Rho
Female Choral	103	.02
Female Instrumental	12	.05
Male Choral	26	.30
Male Instrumental	68	.19

From the Spearman rho analyses, the null hypotheses failed to be rejected: there were no significant relationships between G and R for the 4 teacher type groups. In other words, there were significant differences between the music educators' G and R ranked scores. This finding substantiated the fact that there were at least two possibly different domains of music teaching success being measured in the study: an objective domain as measured by pupil growth and a subjective domain as measured by the ratings of music teaching success. This finding is similar to findings reported in Chapter II by many research studies conducted by A. S. Barr and his associates. This finding led the investigator to additional correlational analyses in order to determine whether or not the way in which the 5 ratings were weighted in order to derive the factor R score criterion could have produced a biased R criterion score.

Correlational Analyses

As shown previously in this chapter, the 5 ratings of music teaching success for the music educator sample were all significantly intercorrelated beyond the .05 level. It was necessary to investigate the intercorrelations among the gain score and the 5 rating criteria. The music educator sample was rank ordered according to G within the three statistically significant pretesting time groups and assigned a single ranked score according to the percentage of subjects within each pretesting group. This prohibited the use of the class mean residual gain score as such within a correlational analysis because this gain score was biased due to pretesting time differences. Therefore, the music educator ranked gain score sample was standardized within each subsample of teacher type. The subjects were assigned a number from 6 to 1 which represented the normal curve of three standard deviations above and below the mean ranked score: the top 3 percent of the sample received a score of 6 while the bottom 3 percent of the sample received a score of 1. With this technique, the class mean residual gain score rank was standardized over the normal bell shaped curve. The 5 rating cards per teacher for each rating category had already been standardized to a scale ranging from 10 to 1 with a standard deviation of 2.00 and a mean of 5.50. These rating scores were used in the correlation analyses. The purpose of standardizing the within groups scores was to reduce variance.

Table 64 shows the intercorrelation matrix which investigated the null hypotheses that there were no significant correlations among the gain and rating criteria and sex beyond the .05 level for a two tailed test.

TABLE 64.--GAIN, RATING, AND SEX CORRELATION MATRIX

N=208	Sex	Gain	Pupil	Principal	Supervisor	Peer	Self
Sex	1.00						
Gain	- .00	1.00					
Pupil	.06	.10	1.00				
Principal	- .13	.03	.34**	1.00			
Supervisor	- .01	.05	.43**	.64**	1.00		
Peer	.02	.09	.45**	.56**	.57	1.00	
Self	- .03	.05	.30**	.19	.13	.19*	1.00

p<.01**, p<.05*

There were no significant correlations between sex and each of the 5 ratings nor between gain and each of the 5 ratings. The null hypotheses failed to be rejected. The preceding table once again demonstrates the finding that there were significant intercorrelations among the 5 ratings with the exception of the self-supervisor correlation, and the null hypotheses beyond the .05 level were rejected. The finding was that while miniscule but positive relationships existed between gain and the 5 rating categories, these were insignificant further substantiating the finding of two possible domains of music teaching success: objective (G) and subjective (R).

Because G and R appeared to be two different domains as indicated by the findings, the 4 criterion variables were used in the final analyses: G, R, G+R, and G-R. In this manner, the relationships between personality and motivational variables and the criterion variables were investigated without empirically deciding which criterion variable was better. Had significant correlations been found between G and R, the final analyses would have concentrated on investigating the relationships to the composite criterion of G+R thus greatly simplifying the study. Since this was not the case, the criterion variables were separated; and because of possible error, G+R was investigated, too.

Table 65 shows the correlation matrix for the teacher type groups and male and female groups for the total derivation sample which was used to develop the regression equations. The music educator cross validation sample (N=41) which represented 20 percent of the sample who were randomly selected from the teacher type groups was not included in the following analysis.

TABLE 65.--CRITERION DATA INTERCORRELATION MATRIX

N=168 Derivation	Rating (R)	Gain (G)	Gain-Rating (G-R)	Gain+Rating (G+R)
Rating (R)	1.00			
Gain (G)	.13	1.00		
Gain-Rating (G-R)	-.67**	.65**	1.00	
Gain+Rating (G+R)	.76**	.75**	-.03	1.00
Male Instrumental N=55				
Rating (R)	1.00			
Gain (G)	.23	1.00		
Gain-Rating (G-R)	-.66**	.59**	1.00	
Gain+Rating (G+R)	.80**	.77**	-.07	1.00
Male Choral N=21				
Rating (R)	1.00			
Gain (G)	.42*	1.00		
Gain-Rating (G-R)	-.56**	.52**	1.00	
Gain+Rating (G+R)	.85**	.84**	-.03	1.00
Total Male N=76				
Rating (R)	1.00			
Gain (G)	.28*	1.00		
Gain-Rating (G-R)	-.63**	.57**	1.00	
Gain+Rating (G+R)	.81**	.79**	-.06	1.00
Female Choral N=82				
Rating (R)	1.00			
Gain (G)	.00	1.00		
Gain-Rating (G-R)	-.71**	.71**	1.00	
Gain+Rating (G+R)	.71**	.71**	.00	1.00
Female Instrumental N=10				
Rating (R)	1.00			
Gain (G)	.15	1.00		
Gain-Rating (G-R)	-.67*	.64*	1.00	
Gain+Rating (G+R)	.77**	.75**	-.03	1.00
Total Female N=92				
Rating (R)	1.00			
Gain (G)	.02	1.00		
Gain-Rating (G-R)	-.70**	.70**	1.00	
Gain+Rating (G+R)	.71**	.71**	-.00	1.00

p<.01**, p<.05*

The above correlation matrices show a significant difference between the male and female groups as far as the relationship between rating and gain was concerned. The male instrumental group's R to G correlation was significant beyond the .10 level which, for the purposes of the study which accepted the .05 level as minimal, was insignificant. However, the male choral group's correlation between R and G was significant beyond the .05 level. The total male derivation sample produced a R to G correlation significant beyond the .02 level for a two tailed test. This finding indicates that for male music educators, there was a relationship between their residual gain score rank and their factorially derived rating score rank. This contradicts the finding of the Spearman's rho analysis which found no significant relationships between the two sets of ranked data. As Dubois mentioned, rho may underestimate the relationship which appeared to be substantiated in the above findings. However, rho was computed by hand while the above correlations treated the ranked data as continuous within the regular Pearson computer formula. Therefore, the latter finding would indicate less possibility for error, and could be due to the fact that males were rated more objectively than females in the study.

The female samples, on the other hand, showed that there were no significant relationships between G and R. These correlations were barely positive in direction indicating that relationships existed but that these were minimal and insignificant. When the male and female groups were combined for the total derivation sample, the female groups which included more subjects contributed in such a way to the G and R correlation that no significant relationship existed.

The G-R to R correlations were all significantly negative beyond the .05 level further finding that two different domains of music teacher accountability existed. All the G-R to G correlations were significant beyond the .05 level which indicated that a gain score construct of music teaching success remained stable even though the G criterion score was reduced by subtracting the R ranked score from it.

The G+R to R and the G+R to G correlations were very close in actual numerical figures and were significant for all subsamples beyond the .01 level. Taking the other correlations into account, the finding was that the R criterion aided the correlation to R and lowered the correlation somewhat to G.

The G+R to G-R correlations for the subsamples and total derivation sample were all right around zero as would be expected with zero order correlations. This finding substantiates the claim that for this sample of music educators, G and R were not significantly related.

Summary

To conclude this section, the four constructed criterion variables of music teaching success identified different domains as indicated by the insignificant correlations for the total derivation sample. Differences in the criterion domain relationships existed between male and female groups of music educators, but these differences were not substantial enough to affect the total combined derivation sample.

The objective measure of music teaching success as identified by the class mean residual gain score technique which derived the G criterion was not related significantly to the subjective measures of music teaching success as identified by the 5 ratings of music teaching success and the factor analytically derived R criterion. Two different domains were evaluated and were treated as such in the following analyses.

THE RELATIONSHIPS AND INTERRELATIONSHIPS AMONG PERSONALITY, MOTIVATIONAL, AND CRITERION VARIABLES

The second subproblem asked the question: what are the relationships and interrelationships of the personality variables measured by the Sixteen Personality Factor Questionnaire and the motivational variables measured by the Motivation Analysis Test to music teaching success? The investigator concentrated his efforts in answering this question using multiple regression techniques.

The first analysis artificially stratified the total music educator sample into 9 success groups: (1) high rating-high gain, (2) high rating-medium gain, (3) high rating-low gain, and so forth to (9) low rating-low gain. Investigated were the relationships among these success groups, sex, and the 36 personality-motivational variables. A significant sex and personality-motivational main effect was found beyond the .0001 level. The MANOVA analysis found that significant sex-personality interactions existed and that personality-teacher type group interactions might occur. This finding precipitated the need to search further in order to identify these interactions as far as their import to the study was concerned.

The next analyses generated the correlations between the personality and motivational variables and success criteria for each teacher type-sex group. Knowing that there were interactions between sex and the personality and motivational variables, these correlational analyses aided in locating these interactions variable by variable, since these showed up as differences between subsamples in the experimental-criterion relationships. Tables 66 and 67 show these correlation matrices.

TABLE 66.--CORRELATION MATRICES - 16 PERSONALITY FACTOR QUESTIONNAIRE VARIABLES

TOTAL MALE N = 76

	A	B	C	E	F	G	H	I	L	M	N	O	Q1	Q2	Q3	Q4
R	-.08	.33	-.03	.11	-.02	.13	.05	.18	.01	.18	-.06	-.08	.08	-.02	.06	.06
G	.05	.16	-.08	.14	-.07	.16	-.02	.04	.04	-.04	.00	.06	.12	.08	.03	.06
G-R	.12	-.15	-.04	.07	-.04	.02	-.06	-.13	.03	-.19	.05	.11	.03	.08	-.02	-.00
G+R	-.02	.31	-.07	.16	-.05	.18	.02	.14	.03	.09	-.04	-.01	.12	.04	.05	.07

TOTAL FEMALE N = 92

	A	B	C	E	F	G	H	I	L	M	N	O	Q1	Q2	Q3	Q4
R	.07	.01	.22	-.08	-.13	.34	.11	.10	-.12	-.19	.12	-.21	-.11	-.07	.34	-.15
G	.10	.06	.11	.14	.06	.09	.11	.08	-.10	.12	-.03	-.20	-.01	-.14	.02	-.08
G-R	.02	.04	-.08	.16	.07	-.18	-.00	-.02	.02	.22	-.11	.00	.07	-.06	-.23	.05
G+R	.12	.05	.23	.04	.02	.30	.16	.12	-.15	-.05	.06	-.28	-.08	-.15	.26	-.16

MALE INSTRUMENTAL N = 55

	A	B	C	E	F	G	H	I	L	M	N	O	Q1	Q2	Q3	Q4
R	-.07	.26	-.05	.05	-.02	.05	.12	.19	.02	.13	-.10	-.09	.10	-.06	.10	.04
G	.10	.05	-.12	.21	-.07	.08	.05	.09	.11	-.11	.00	.16	.15	.02	-.00	.11
G-R	.13	-.17	-.05	.12	-.04	.02	-.06	-.09	.06	-.20	.09	.20	.04	.07	-.08	.05
G+R	.02	.20	-.11	.16	-.05	.08	.11	.18	.08	.02	-.07	.04	.16	-.03	[illegible]	.09

MALE CHORAL N = 21

	A	B	C	E	F	G	H	I	L	M	N	O	Q1	Q2	Q3	Q4
R	-.13	.55	.05	.24	-.02	.34	-.17	.18	-.02	.29	.06	-.05	.03	.08	-.08	.14
G	-.05	.47	.05	-.00	-.07	.38	-.23	-.11	-.09	.10	-.01	-.20	.05	.22	.14	-.17
G-R	.08	-.09	-.01	-.23	-.05	.04	-.05	-.27	-.07	-.17	-.06	-.14	.02	.12	.19	-.29
G+R	-.11	.60	.06	.15	-.05	.43	-.23	.04	-.07	.23	.03	-.15	.05	.18	.05	-.01

FEMALE CHORAL N = 82

	A	B	C	E	F	G	H	I	L	M	N	O	Q1	Q2	Q3	Q4
R	.02	.05	.16	-.08	-.01	.33	.09	.14	-.02	-.12	.09	-.13	-.07	.00	.33	-.08
G	.14	.10	.08	.12	.07	.07	.14	.14	-.12	.17	-.04	-.20	.05	-.19	-.05	-.06
G-R	.08	.04	-.05	.14	.06	-.18	.04	.01	-.07	.21	-.10	-.06	.08	-.13	-.27	.01
G+R	.12	.10	.17	.03	.14	.28	.16	.20	-.09	.03	.04	-.23	-.01	-.13	.20	-.10

FEMALE INSTRUMENTAL N = 10

	A	B	C	E	F	G	H	I	L	M	N	O	Q1	Q2	Q3	Q4
R	.25	-.31	.69	.00	-.22	.38	.30	-.29	-.76	-.64	.40	-.72	-.53	-.47	.44	-.67
G	-.16	-.28	.32	.35	-.03	.25	-.25	-.56	.05	-.27	.06	-.18	-.65	.12	.68	-.20
G-R	-.32	.03	-.30	.26	.15	-.11	-.43	-.20	.63	.30	-.26	.43	-.08	.45	.17	.37
G+R	.07	-.38	.67	.23	-.17	.42	.04	-.56	-.47	-.61	.31	-.60	-.77	-.24	.74	-.58

BEST COPY AVAILABLE

TABLE 66.--CONTINUED

SIGNIFICANCE LEVELS FOR A TWO TAILED TEST

N	.05 level	.01 level	N	.05 level	.01 level
76	.22	.29	21	.42	.54
92	.21	.27	82	.22	.28
55	.26	.34	10	.58	.71

TABLE 67.--CORRELATION MATRICES - MOTIVATION ANALYSIS TEST VARIABLES

ALL MALE N = 76

	U-CA	U-HO	U-FR	U-NA	U-SE	U-SS	U-MA	U-PG	U-AS	U-SW
R	-.07	-.10	.07	.07	-.03	-.01	.10	-.06	-.22	.11
G	-.07	.13	-.19	-.02	.01	-.13	.05	-.10	-.08	-.04
G-R	.00	.19	-.21	-.07	.04	-.10	-.04	-.03	.12	-.13
G+R	-.09	.01	-.07	/04	-.01	-.09	.10	-.10	-.20	.05

	I-CA	I-HO	I-FR	I-NA	I-SE	I-SS	I-MA	I-PG	I-AS	I-SW
R	-.00	.13	.05	-.01	-.00	.09	-.02	-.09	-.02	.09
G	.13	.08	-.00	-.15	.08	-.12	.11	.18	-.02	-.10
G-R	.11	-.05	-.04	-.15	.07	-.18	.11	.23	-.00	-.16
G+R	.08	.13	.03	-.10	.05	-.02	.06	.05	-.02	-.00

ALL FEMALE N = 92

	U-CA	U-HO	U-FR	U-NA	U-SE	U-SS	U-MA	U-PG	U-AS	U-SW
R	.11	.05	-.05	[illegible]	-.09	.03	-.02	-.09	-.05	-.00
G	-.13	-.17	.16	.00	.06	.32	-.06	.02	-.14	.10
G-R	-.17	-.15	.15	-.06	.11	.21	-.02	.07	-.07	.08
G+R	-.01	-.08	.08	.06	-.02	.25	-.05	-.05	-.14	.07

	I-CA	I-HO	I-FR	I-NA	I-SE	I-SS	I-MA	I-PG	I-AS	I-SW
R	-.16	.00	-.08	.15	-.07	-.02	.01	-.02	-.07	.15
G	.13	.00	.08	.02	.01	.10	-.05	-.05	-.09	.03
G-R	.20	.00	.12	-.10	.05	.08	-.04	-.03	.02	-.09
G+R	-.02	.00	-.00	.12	-.04	.06	-.03	-.05	-.12	.12

MALE INSTRUMENTAL N = 55

	U-CA	U-HO	U-FR	U-NA	U-SE	U-SS	U-MA	U-PG	U-AS	U-SW
R	-.19	-.21	.22	.14	-.22	-.01	-.01	.12	-.29	.31
G	-.05	.09	-.15	.04	-.09	-.13	.07	-.01	-.13	.04
G-R	.12	.25	-.30	-.09	.11	-.09	.06	-.11	.14	-.23
G+R	-.15	-.08	.05	.11	-.21	-.09	.05	.08	-.27	.23

	I-CA	I-HO	I-FR	I-NA	I-SE	I-SS	I-MA	I-PG	I-AS	I-SW
R	.03	-.02	.05	.03	.01	.05	.02	-.09	-.04	.10
G	.13	-.09	-.05	-.27	.05	-.04	.14	.08	-.09	-.10
G-R	.08	-.05	-.08	-.23	.03	-.07	.10	.13	-.04	-.16
G+R	.10	-.07	.00	-.15	.03	.01	.10	-.01	-.09	.01

MALE CHORAL N = 21

	U-CA	U-HO	U-FR	U-NA	U-SE	U-SS	U-MA	U-PG	U-AS	U-SW
R	.31	.16	-.30	-.12	.52	-.00	.35	-.48	.02	-.41
G	-.17	.20	-.28	-.15	.31	-.14	.02	-.31	.07	-.26
G-R	-.45	.03	.03	-.03	-.21	-.13	-.32	.17	.05	.15
G+R	.09	.21	-.34	-.16	.49	-.08	.22	-.47	.05	-.40

	I-CA	I-HO	I-FR	I-NA	I-SE	I-SS	I-MA	I-PG	I-AS	I-SW
R	-.08	.53	.04	-.11	-.04	.17	-.09	-.10	.06	.07
G	.12	.48	.14	.16	.22	-.34	.05	.42	.19	-.09
G-R	.18	-.06	.09	.25	.24	-.48	.14	.48	.11	-.15
G+R	.02	.59	.10	.02	.11	-.10	-.02	.19	.15	-.01

SIGNIFICANCE LEVELS FOR A TWO TAILED TEST

N	.05 level	.01 level	N	.05 level	.01 level
76	.22	.29	21	.42	.54
92	.21	.27	82	.22	.28
55	.26	.34	10	.58	.71

TABLE 67.--CONTINUED

FEMALE CHORAL N = 82

	U-CA	U-HO	U-FR	U-NA	U-SE	U-SS	U-MA	U-PG	U-AS	U-SW
R	.12	-.01	-.05	.08	-.11	.07	.03	-.01	-.12	.01
G	-.11	-.17	.12	-.02	.09	.30	-.03	-.00	-.13	.08
G-R	-.17	-.11	.12	-.07	.14	.16	-.04	.00	-.01	.05
G+R	.01	-.13	.05	.04	-.02	.27	.00	-.01	-.18	.06
	I-CA	I-HO	I-FR	I-NA	I-SE	I-SS	I-MA	I-PG	I-AS	I-SW
R	-.10	-.03	-.10	.14	-.08	-.03	.03	-.02	-.07	.19
G	.08	-.02	.07	.03	.02	.20	-.10	-.15	-.09	.03
G-R	.13	.00	.12	-.08	.07	.16	-.09	-.10	-.02	-.11
G+R	-.01	-.03	-.02	.11	-.04	.12	-.04	-.12	-.11	.15

FEMALE INSTRUMENTAL N = 10

	U-CA	U-HO	U-FR	U-NA	U-SE	U-SS	U-MA	U-PG	U-AS	U-SW
R	-.03	.56	-.17	.17	.07	-.42	-.37	-.51	.50	-.04
G	-.24	-.14	.57	.21	-.21	.50	-.26	.16	-.22	.28
G-R	.16	-.54	.56	.03	-.21	.70	.10	.51	-.56	.25
G+R	-.18	.28	.25	.25	-.09	.04	-.41	-.23	.19	.16
	I-CA	I-HO	I-FR	I-NA	I-SE	I-SS	I-MA	I-PG	I-AS	I-SW
R	-.60	.17	.04	.29	-.01	.07	-.22	.01	-.16	-.29
G	.55	.17	.17	-.03	-.16	-.71	.23	.79	-.08	-.04
G-R	.88	-.00	.10	-.24	-.12	-.59	.35	.58	.07	.19
G+R	-.04	.22	.14	.18	-.11	-.41	.00	.52	-.16	-.22

The above correlation matrices show some significant relationships between personality and motivational variables and music teaching success criteria. The main purpose of these correlations was not to select variables for the ensuing multiple regression equations. All of the variables were used although many studies have used significant correlations initially to reduce the number of variables for regression analyses. By using this approach, one might exclude certain variables which were not significantly related but yet contribute to the power of the equation as suppressor variables. Furthermore, simply because significant correlations exist between experimental variables and criteria do not mean that these variables will be useful in investigating experimental and criterion variables' relationships using multiple regression techniques.

The interactions were located among all four criteria as far as the correlations between the experimental variables and the criteria differed for the four teacher type sex groups. While there were sometimes more pronounced differences for an experimental variable between the groups on one of the 4 criteria, the selection of the following cross product, moderating variables, was made from differences in the correlations for the 4 criteria. Only one type interaction appeared sufficient enough to warrant inclusion as a variable in the regression equations that used the 16PF variables. This was Factor O (self-assured versus apprehensive). Therefore, type was included as a variable along with Factor O times Type as a moderating variable.

These cross product terms which were included in the regression equations as moderating variables accounting for the sex interactions were formulated by multiplying the experimental variable times sex. While some may consider this technique merely palliative, it was the only statistical technique available which allowed the total music educator derivation sample to be used for investigating relationships to music teaching success.

Eight interactions were used for constructing the sex moderating variables. These were 16PF variables E, G, M, and N, and MAT variables HO-U, SS-U, SS-I, and PG-I. The correlations indicated that the male music educator tended to be more accommodating and the female more stubborn (Factor E); that the male was more conscientious while the female more expedient (Factor G); that the male was more imaginative while the female more practical (Factor M); that the male was more shrewd, astute and socially polished and the female more forthright and unpretentious (Factor N); that the male had more unsatisfied needs directed towards the home than the female (HO-U); that the male was less social than the female who tended towards social approval (SS-U and SS-I); and that the male tended to be more pugnacious and defensive than the female (PG-I). The overall pattern from the correlational analyses suggested that the male music educator was different from his female counterpart in that he was less social and more attached to the home, lacked drive in the self sentiment area, tended to be more accommodating and submissive, and was conscientious and practical while shrewdly defending his position or career. The female music educator, on the other hand, tended to be driven more towards social approval, needed less satisfaction from the home, was more stubborn and tended to be more expedient and less defensive but more imaginative and forthright. These differences appeared to be of sufficient importance to treat them statistically as cross product terms and bring them into the regression equations as sex-personality moderating variables. Sex as a separate experimental variable was also introduced in the equations.

Phase One Findings

The main purpose of the study was to investigate the relationships and interrelationships among personality, motivational, and criterion variables. Multiple regression techniques predict the success criteria from the personality and motivational variables alone. These predictions provide useful information about the relationships and interrelationships.

Previous to investigating these relationships, it was necessary to identify the possible interactions among the experimental variables and the criterion and sex and type variables. One type and 8 sex interaction variables were identified from the correlations of the experimental variables to the criterion variables for the male and female music educator groups. With these analyses completed, the multiple regression analyses were undertaken.

From the final sample of 209 subjects, a derivation sample of 168 music educators and a cross validation sample of 41 subjects were stratified. Stepwise multiple regression equations were formulated to predict music teaching success for each of the 4 criterion variables. The stepwise computer program that was used in the study allowed for both the inclusion of progressively significant variables and the exclusion of variables which no longer contributed significantly to the multiple R. The 16PF variables were used separately along with the moderating variables in the first analyses because these variables were psychometrically different from the MAT motivation variables. The 16PF variables identify the basic temperament-trait domain of psychometric theory while the MAT variables reflect the dynamic domain of motivations, drives, sentiments, feelings, and ergs. Therefore, separate multiple regression equations were formulated for the sixteen 16PF and 5 moderating variables to music teaching success criteria and for the 20 MAT and 4 moderating variables to music teaching success criteria.

The null hypotheses that no significant relationships existed between the experimental personality and motivational variables and criterion variables were investigated beyond the .05 level, the significant relationships, when tested on an independent sample, would not cross validate significantly beyond the .05 level were also investigated. Table 68 shows the 16PF variables.

TABLE 68.--SIXTEEN PERSONALITY FACTOR QUESTIONNAIRE VARIABLES

Factor A	Sizothymia Stiff	Affectothymia Participating
Factor B	Low Intelligence Dull	High Intelligence Bright
Factor C	Low Ego Strength Easily Upset	High Ego Strength Calm
Factor E	Submissiveness Accommodating	Dominance Stubborn
Factor F	Desurgency Serious	Surgency Enthusiastic
Factor G	Weak Superego Expedient	Strong Superego Conscientious
Factor H	Threctia Shy	Parmia Venturesome
Factor I	Harria Realistic	Premsia Sensitive
Factor L	Alaxia Trusting	Protension Suspicious
Factor M	Praxernia Practical	Autia Imaginative
Factor N	Artlessness Forthright	Shrewdness Astute
Factor O	Untroubled Adequacy Self Assured	Guilt Proneness Apprehensive
Factor Q1	Conservative Conservative	Radicalism Experimenting
Factor Q2	Group Adherence Group Dependent	Self Sufficiency Self Sufficient
Factor Q3	Low Self Sentiment Integration Self Conflict	High Self Sentiment Integration Controlled, Socially Precise
Factor Q4	Low Ergic Tension Relaxed	High Ergic Tension Tense

Table 69 shows the stepwise multiple regression equation which was most significant for investigating the relationships between personality variables and the G criterion. This equation produced a multiple R = .26 which was significant beyond the .05 level. It is interesting to note that none of the moderating variables contributed to the equation.

When tried out on the cross validation sample, the equation was not substantiated; $r_c = -.08$. In order to assure that the invalid cross validation r was not a computer error, the programs were run twice and the same results occured. Therefore, while the null hypothesis which investigated the relationship between personality and G could be rejected initially, upon cross validation, the relationships were not substantiated, and the null hypotheses could not be rejected. The finding is that the G criterion was not related to the 16PF variables.

Table 70 shows the multiple regression equation which best investigated the relationships between the personality variables and the R criterion. This equation included two sex moderating variables as well as sex alone as additional experimental variables along with the personality variables. The equation was significant beyond the .05 level and produced a respectible multiple R of .42 which accounted for 17 percent of the variance. The initial finding was that the null hypothesis could be rejected.

When the equation was applied to the cross validation sample, no significant results were found. Because of the size of the negative cross validation correlation (r = -.23), the computer programs were run twice with like results. The null hypothesis that the equation would not be significant on cross validation failed to be rejected. The finding was that the 16PF personality variables were not related to the R criterion.

THIS PAGE WAS MISSING FROM THE DOCUMENT THAT WAS SUBMITTED TO ERIC DOCUMENT REPRODUCTION SERVICE.

Table 71 contains the multiple regression equation which best demonstrated the relationship between personality and the G+R criterion. Two sex moderating and one type moderating variables entered the equation which was significant beyond the .05 level. The multiple correlation coefficient was respectable (R=.39) which indicated that 15 percent of the variance in the criterion could be accounted for by the personality variables alone. The null hypothesis was rejected.

The equation did not cross validate significantly indicating that the significant relationships were not stable when applied to a like sample of subjects. Because the cross validation coefficient was negative (r= -.13), this equation, too, was run twice with the same results. The null hypothesis for cross validation failed to be rejected. The finding is that the personality variables measured by the 16PF test were not related to the G+R criterion.

Table 72 (Stepwise Program) shows the complete stepwise multiple regression program for each step. The first variable to enter the equation was Q3 and the last was teacher TYPE. None of the 20 steps of the equation was significant.

Table 72 (16PF Multiple Regression Analysis) shows the equation for investigating the relationships between personality variables and the G-R criterion. This equation was not significant, and the null hypothesis could not be rejected: the 16PF personality variables were not related to the G-R criterion. Since the equation was not significant, no cross validation was attempted.

Cross validation was important to the study. Even though 3 of the equations were significant, the 16PF personality variables were related to the G, R, and G+R criteria, these relationships were not substantiated on the cross validation sample. The finding was that the basic temperament-trait domain as measured by the 16PF test was not related to music teaching success in terms of

TABLE 71.--16PF MULTIPLE REGRESSION ANALYSIS - GAIN + RATING CRITERION

Variables	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
B	.19	.07	.20	.19	.08
C	-.04	.04	-.08	-.10	.10
G	.12	.04	.23	.24	.08
I	.07	.05	.13	.14	.09
O	-.14	.06	-.17	-.34	.15
O X TYPE	.03	.03	.07	.14	.15
M X S	-.03	.02	-.12	-.18	.12
E X S	.04	.02	.15	.21	.11

Constant = -2.86

R = .39 R^2 = .15 R' = .33 DF = 8/159 r_c = -.13 F = 3.49*

p<.05*

TABLE 72.--16PF MULTIPLE REGRESSION ANALYSIS - GAIN - RATING CRITERION

Variable	F Ratio	R	R^2	R'
Q3	2.76	.13	.02	.13
G X S	2.29	.16	.03	.15
E X S	1.81	.18	.03	.14
G	1.63	.20	.04	.14
SEX	1.44	.21	.04	.14
B	1.42	.22	.05	.14
H	1.40	.24	.06	.15
M X S	1.31	.25	.06	.14
M	1.60	.29	.08	.19
E	1.53	.30	.09	.19
A	1.47	.31	.09	.19
I	1.43	.32	.10	.19
F	1.36	.32	.10	.18
C	1.26	.32	.10	.17
L	1.19	.32	.10	.15
Q4	1.10	.32	.10	.13
O X TYPE	1.04	.32	.11	.10
N	.97	.32	.11	.06
N X S	.95	.33	.11	.04
TYPE	.90	.33	.11	.07

The above represents 20 steps or variables entering the stepwise program in their order of entrance: Q3 was first and TYPE was last. None of the steps was significant beyond the .05 level as represented by the F ratio.

G, R, G+R, and G-R. This finding differs significantly from the previous literature which did not use a multivariate approach but rather used correlation, t-test, or nonparametric statistical analyses. While some may conclude from the above that these findings may be in doubt, the importance of cross validation cannot be overemphasized.

> . . . a careful researcher would not put much stock in a regression equation until he has cross-validated it on a sample other than the one on which it was based . . . Therefore, it has become customary to set aside a certain fraction of the total original sample for the purpose of cross validation (Tatsuoka 1969:26-27).

The Motivation Analysis Test (MAT) is an objective questionnaire which measures the dynamic psychometric domain of drives, sentiments, feelings, and values: in essence, it measures the strengths of a person's motivations on 10 variables in two ways thus providing a total of 20 motivational experimental variables. First, the Unintegrated Motive Strengths are measured. These reflect the unsatisfied need and aspiration levels and are, for the most part, unconscious drives (ergs) and sentiments (feelings) rising out of the dynamic energy domain.

Second, the Integrated Motive Strengths are measured. These reflect the portion of the total dynamic energy that has arisen from successful fulfillment of these needs and aspirations. Integrated drives and sentiments reflect a type of habit strength which was consiously built from a person's previously successful experiences in these measured areas. Table 73 shows the motivational factors.

TABLE 73.--MOTIVATION ANALYSIS TEST VARIABLES

Career (CA)	A measure of a person's strength of interest in his career, his feelings or sentiments to his chosen career.
Home-Parental (HO)	A measure of a person's strength of interest in his home and familial surroundings, his feelings or sentiments to the home.
Fear (FR)	The fear or escape erg or drive refer to security. A measure of a person's interest in avoiding illness, accident, loss of financial security, military threats, and death. It is called an escape erg with fear as the accompanying emotion.
Narcism-Comfort (NA)	This is the strength of a person's drive toward a sensual indulgence of all kinds such as food and smoking, to ease, self love, and avoidance of onerous duties. NA refers to Freud's narcistic, sex component.
Superego (SE)	A measure of a person's strength of interest in the way he feels about getting things accomplished. It is similar to Factor G - 16PF in that a strong or high superego sentiment relates to conscientiousness while a low interest in superego relates to expediency.
Self Sentiment (SS)	A measure of a person's interest in social approval and his degree of concern about his general reputation.
Mating (MA)	A measure of a person's strength of interest in the opposite sex, romantic literature, courting behavior, beauty, and sensuality. Mating is the sex drive.
Pugnacity-Sadism (PG)	This is a measure of a person's strength of his drive to attack, damage, inflict pain, and destroy. The accompanying emotion is rage in the extreme.
Assertiveness (AS)	A measure of a person's strength of interest or drive towards status. It is one of the roots of ambition and is similar to pride but best called vanity or assertiveness. It shows itself in striving for admiration, wishing to excel in competition, aiming to be handsomely dressed, and moving in high circles.
Sweetheart-Spouse (SW)	A measure of the strength of a person's interest in his wife. The strength of his feeling or sentiment to the opposite sex.

Table 74 shows the multiple regression equation which best investigated the relationships between the motivational variables and G. All 20 MAT variables plus the 4 moderating variables as well as the sex and teacher type variables were included in the stepwise multiple regression program. The G equation included 11 variables which were significant contributors. The equations, which produced a multiple correlation coefficient of .35 and accounted for 12 percent of the variance, was significant beyond the .05 level.

The cross validation coefficient (r=.40) was significant beyond the .01 level indicating that the equation was somewhat stable and that motivational variables were related to music teaching success in terms of gain (G). The null hypotheses that no significant relationships existed and cross validated were rejected.

It is interesting to note that the equation as corrected for shrinkage produced a multiple correlation coefficient of .25; yet, the equation was actually at a higher level of significance and a higher numerical coefficient for the cross validation sample. In most cases, one might expect the multiple R to be reduced but such was not the case with this equation. Therefore, the cross validation r indicated that 16 percent of the variance in the gain criterion could be accounted for from motivational variables alone. This is a significant finding.

The equation which highlighted the relationships between motivation and G success, indicate that the following variables contributed significantly to the equation: CA-, HO+, SS-, AS-, HO+, and PG+ along with the moderating variables which showed that differences in the male and female subjects affected the nature of the relationships. It may appear confusing to the reader to see Career as both plus and minus, positive and negative. This indicated that the feeling towards career was a conscious and not an unconscious sentiment since the U score was negative and the I score was positive.

Success in music teaching, as defined by the G criterion, is related to greater attachment to the home (HO+), lower need for social approval (SS-),

TABLE 74.--MULTIPLE REGRESSION ANALYSIS MAT VARIABLES - GAIN CRITERION

Variable	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
CA-U	-.06	.05	-.09	-.09	.08
HO U	.22	.12	.15	.40	.22
SS U	-.26	.12	-.17	-.41	.19
AS U	-.04	.04	-.07	-.07	.08
CA I	.04	.04	.08	.03	.08
HO I	.03	.05	.06	.06	.08
PG I	.19	.10	.15	.41	.22
PG I X S	-.10	.06	-.13	-.41	.24
HU X S	-.15	.07	-.17	-.61	.28
SU X S	.21	.07	.23	.84	.28
SI X S	.02	.03	.05	.09	.13

Constant = -.56

$R = .35$ $R^2 = .12$ $R' = .25$ DF = 11/156 $r_c = .40$** F = 1.92*

$p<.01$**, $p<.05$*

lower need for status (AS-) which suggests that the music educator is less sociable. He is generally satisfied with his chosen career which was not causing him any conflicts as indicated by the pattern of the U and I scores (CA-U,CA+I), and exhibits pugnacious behavior which in its Integrated expression here (PG+I) indicates long standing defensiveness and hostility. The overall behavior pattern suggests a loner, a person who does his own thing and does not necessarily care what his peers think about him. PG+I indicates that the music educator who is successful in gain tends to defend his position even to the point of hostility. This could be interpreted to mean that since the good music educator tends to defend his music program and fight for its existence; in the process of so doing, he becomes somewhat hostile and defensive. Therefore, the successful music educator tends to be less sociable and regards his home as important. An interesting observation is that the successful music educator appears lower in need for admiration and competition (AS-U).

Table 75 shows the complete stepwise output for investigating the relationships between the motivational variables and the R criterion. None of the steps produced equations significant beyond the .05 level. The null hypothesis could not be rejected. The finding was that no relationships appeared to exist that were significant beyond the .05 level.

Table 76 shows the stepwise output for investigating the relationships between the motivational variables and the G+R criterion. None of the steps was significant beyond the .05 level, and the null hypothesis could not be rejected. There were no significant relationships between the motivational variables and G+R.

Table 77 shows the best equation for investigating the relationships between the motivational variables and the discrepancy criterion (G-R). This equation was significant beyond the .05 level for both the derivation and cross validation samples thus rejecting the null hypotheses that no significant

TABLE 75.--MULTIPLE REGRESSION ANALYSIS - MAT VARIABLES - R CRITERION

Variables	F Ratio	R	R^2	R'
AS-U	2.74	.13	.02	.13
SW-I	2.56	.17	.03	.16
CA-I	2.64	.21	.05	.19
NA-I	2.32	.23	.05	.19
AS-I	1.99	.24	.06	.19
PG-U	1.77	.25	.06	.18
SE-U	1.64	.26	.07	.18
HO-I	1.49	.26	.07	.17
SE-I	1.37	.27	.07	.16
FR-U	1.26	.27	.07	.15
MA-U	1.15	.27	.08	.13
MA-I	1.06	.28	.08	.10
CA-U	.98	.28	.08	.07
NA-U	.92	.28	.08	.03
HO-U X S	.85	.28	.08	.08
HO-U	.80	.28	.08	.11
SEX	.78	.28	.08	.13
SS-I X S	.73	.29	.08	.15
SS-I	.70	.29	.08	.17
SEX (removed)	.74	.29	.08	.19
SW-U	.70	.29	.08	.17
SS-U	.66	.29	.08	.19
SS-U X S	.64	.29	.08	.20
SEX	.61	.29	.08	.22
TYPE	.58	.29	.08	.23
FR-I	.55	.29	.08	.25
PG-I	.52	.29	.08	.26

Complete stepwise output: first variable entering the equation was AS-U and last variable was PG-I. None of the steps were significant beyond the .05 level.

TABLE 76.--MULTIPLE REGRESSION ANALYSIS - MAT VARIABLES - G+R CRITERION

Variables	F Ratio	R	R^2	R'
AS-U	4.47	.16	.13	.16
FC-U	2.75	.18	.03	.16
AS-I	2.06	.19	.14	.16
SS-U X S	1.68	.20	.04	.15
SEX	1.51	.21	.04	.14
SS-U	1.62	.24	.06	.17
HO-U	1.45	.24	.06	.16
SE-U	1.32	.25	.06	.15
HO-U X S	1.21	.25	.06	.13
CA-I	1.11	.26	.07	.11
SW-I	1.04	.26	.07	.09
CA-U	.96	.26	.07	.06
PG-I	.89	.26	.07	.04
PG-I X S	.86	.27	.07	.07
FR-U	.81	.27	.07	.10
SS-I	.76	.27	.07	.13
TYPE	.72	.28	.07	.15
HO-U	.68	.28	.08	.17
MA-U	.65	.28	.08	.19
NA-I	.61	.28	.08	.20
MA-I	.58	.28	.08	.22
SE-I	.55	.28	.08	.23
SS-I X S	.53	.28	.08	.25
FR-I	.50	.28	.08	.26
SW-U	.48	.28	.08	.28

Complete stepwise output: first variable entering the equation was AS-U and last variable was SW-U. None of the steps were significant beyond the .05 level.

TABLE 77.--MULTIPLE REGRESSION ANALYSIS - MAT VARIABLES - G-R CRITERION

Variables	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
CA-U	-.08	.07	-.09	-.09	.08
HO-U	.36	.14	.19	.48	.20
NA-U	-.04	.05	-.07	-.07	.08
SS-U	-.24	.14	-.13	-.28	.17
CA-I	.10	.05	.15	.15	.08
NA-I	-.06	.05	-.09	-.09	.08
SW-I	-.10	.05	-.15	-.15	.08
HU X S	-.23	.08	-.21	-.68	.25
SU X S	.19	.08	.18	.56	.24

Constant = .77

R = .34 R^2 = .12 R' = .27 DF = 9/158 F = 2.29*

p<.05*

relationships existed between the motivational variables and G-R.

The above equation may be the most significant finding of the study because it indicated that motivational variables help to explain the discrepancy between subjective and objective evaluations of a music educator's performance. Many of the same variables which were related to G were also related to G-R but a slightly different pattern was evident.

The music educator whose gain score was better than his rating score appeared lower in need for social approval (SS-U) and more attached to the home (HO+U) but lower in need for career satisfaction (CA-U). He had less interest in sensual indulgences and the avoidance of onerous duties (NA-I, NA-U) and appeared to be satisfied with his mate and had less interest in the opposite sex (SW-I) which balanced his strong attachment to the home. In this equation, only two of the moderating variables contributed (HO-U and SS+U) indicating that the sexes differed significantly in their attachment to home and need for social approval.

The G equation included the pugnacity-sadism variable which might be better called a hostility variable. This was a positive habit strength indicating that the music educator had received satisfaction in the past from behaving this way. Here the interpretation is best centered around defensiveness. Since the music educator must defend his program and literally fight for its existence in many cases, it makes sense that this variable would enter the equation. Defensiveness ties in with the loner classification.

Motivational variables as measured by the Motivation Analysis Test were related significantly to music teaching success in terms of pupil growth (G) and the discrepancy criterion (G-R) but not to ratings (R) nor the composite criterion (G+R). The temperament trait variables measured by the Sixteen Personality Factor Questionnaire did not produce any significant relationships to the 4 music teaching success criteria that remained stable on cross validation.

The above findings must be viewed cautiously, but just how cautiously is naturally open to question. The first section of this chapter discussed in considerable detail the music teaching success criteria and the many pitfalls within the criterion domain. These alone suggest cautious interpretations of these findings. The gain score construct (G) was carefully statistically refined. In practice, most public school systems could not or would not take the time to refine the objective pupil music achievement data in this manner. The derived pupil class mean residual gain score, however, was quite similar to a gain score that might have been formulated using analysis of covariance techniques. Most large school districts have the facilities to perform ANCOVA analyses.

The gain score did not adequately control the regression effects, but it was open to doubt in the study whether or not these were in fact testing biases due to regression or due to teacher effects. If these were due to teacher effects, the study was more valid in this domain because it showed a teacher's ability to produce gain (G) and to select appropriate evaluative measuring instruments, e.g. music achievement subtests. The finding that the G criterion was valid probably should be viewed somewhat cautiously.

The ratings gathered in the study were empirically valid. The analyses showed 3 possible domains within the 5 ratings for the music educator sample, and these were weighted appropriately through factor analytic techniques to derive the R criterion. The rating instrument was valid and discriminated among the music educators in the sample. While rating was not significant as a success criterion does not imply that ratings are invalid. Rather, the ratings mean something else and were tapping information not necessarily related to the experimental variables used in the study as investigated using stepwise multiple regression techniques.

BEST COPY AVAILABLE

Because R was not related to the personality and motivational variables, the music student teacher cross validation sample was not used to test the two significant equations. Furthermore, the validity of the music student teacher rating data was in doubt. The fact that significant motivational relationships existed between the G and G-R criteria was a major finding.

Had the analyses stopped at the point of the significant and well cross validated regression equations, the differences among teacher type groups would have not beeen investigated. Therefore, the above findings may be viewed as Phase One findings at which time it was decided by the investigator to procede with teacher type regression analyses.

Phase Two Findings

As a result of the Phase One findings of the study, it was determined that sex and teacher type were interacting with the personality and motivational variables in fairly powerful ways. The net effect of such interactions was to lower the significance of the relationships investigated with the combined sample, hence, the relatively low but significant multiple R's.

While the effect of teacher type was not predetermined to be a research variable, it was apparent in Phase One Findings. This was unexpected because the literature led the investigator to be alert for possible sex interactions and to methods for controlling these interactions. The correlation matrices hinted at possible teacher type subgroups interactions; yet, only one for the 16PF Test was included as an additional moderator in the Phase One equations. With this in mind, Phase Two of the analyses attempted to find differences in regression equations investigating the relationships among the personality, motivational, and success variables for teacher type groups and to see if the multiple correlation coefficients could be improved significantly.

Separate stepwise multiple regression equations were formulated for each teacher type-sex group for both the 16PF and MAT variables to investigate the relationships of personality and motivational variables to music teaching success in terms of the 4 criterion variables. The total sample (N=209) was used because the derivation sample would have become too small when stratified into these subgroups, and there would have been too few subjects in each group. As it was, there were too few subjects in the male choral and female instrumental groups to make any meaningful interpretations from the data.

The music educator cross validation sample could not be used for the above reasons. However, the music student teacher sample was used as an attempt to cross validate the equations for the R criterion. Furthermore, the subgroups were used primarily to cross validate equations for curiosity reasons. In other words, the equations developed on the male instrumental group were cross validated to the male choral, female instrumental, and female choral groups. Heuristically, this approach made sense even though the samples were small and could not be cross validated adequately.

The null hypotheses investigated were that by partitioning the sample into teacher type-sex groups no significant relationships existed between personality and motivational variables and the 4 criterion variables (G, R, G+R, and G-R) beyond the .05 level. Furthermore, the null hypotheses that no significant cross validations would occur were also tested.

MALE INSTRUMENTAL

Table 78 shows the best multiple regression equation for investigating the relationship between the 16PF personality variables and the R criterion for the male instrumental group. Only two variables contributed to this equation (Factor B and Factor H) which indicated that to some extent intelligence and venturesomeness were related to ratings of music teaching success. The multiple R was lower than for the total group (R=.30) indicating that approximately 9 percent of the variance in the R criterion could be accounted for by the personality variables alone.

TABLE 78.--MULTIPLE REGRESSION ANALYSIS - MALE INSTRUMENTAL - 16PF - R CRITERION

Variables	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
B	.14	.07	.24	.24	.12
H	.03	.02	.17	.17	.12

Constant = -1.71

R = .30 R^2 = .09 R' = .27 F = 3.17*

p<.05*

CROSS VALIDATION

STUDENT TEACHER	
Female Instrumental	-.09
Male Instrumental	-.20
Female Choral	.11
Male Choral	.46
MUSIC EDUCATOR	
Male Choral	.25
Female Choral	.10
Female Instrumental	-.05

The cross validation to the music student teacher male instrumental group was negative while it was positive to the music educator male choral and female choral groups and negative to the female instrumental group. The null hypothesis was rejected; but because the equation failed to cross validate and because it did not account for very much variance in the criterion, the finding was that personality variables had some relationships to music teaching success in terms of R.

Tables 79 through 81 show the stepwise programs for investigating the relationships between the personality variables and the G, G+R, and G-R criteria for the male instrumental group. None of the equations was significant beyond the .05 level, and the null hypotheses were not rejected. The finding was that personality variables appeared to have no significant relationships to G, G+R, and G-R as far as the male instrumental music educator sample was concerned.

Table 82 shows the multiple regression equation for investigating the relationships between the motivational variables and the R criterion for the male instrumental group. The equation was significant beyond the .05 level producing a coefficient of $R=.52$ which indicated that 27 percent of the variance in the R criterion was accounted for by the motivational variables alone. This was a respectable figure, and the null hypothesis was rejected. The finding was that the significance of the relationships improved between the MAT variables and the R criterion for the subgroup of male instrumental music educators. Motivation was related to music teaching success in terms of R.

The cross validation of the equation to the male instrumental music student teacher group produced a negative relationship ($r= -.05$) indicating that these two groups were either rated differently, had varying reliabilities and validities for their R criteria, or that the relationships were not stable due to varying motivational characteristics for the groups. The null hypothesis for cross validation was not rejected.

TABLE 79.--MULTIPLE REGRESSION ANALYSIS - MALE INSTRUMENTAL - 16PF - G CRITERION

Step	Variable	F Ratio	R	R^2	R'	
1	G	2.14	.18	.13	.03	.98
2	Q1	1.92	.23	.06	.04	.98
3	O	1.68	.27	.07	.04	.97
4	M	1.42	.29	.08	.04	.98
5	E	1.21	.30	.09	.03	.98
6	F	1.15	.32	.10	.03	.98
7	A	1.08	.33	.11	.02	.98
8	B	.99	.34	.12	.02	.99
9	L	.92	.35	.12	.01	.99
10	Q3	.84	.36	.13	-.01	1.00
11	N	.77	.36	.13	-.02	1.01
12	Q2	.70	.36	.13	-.04	1.01
13	I	.64	.37	.13	-.06	1.02
14	Q4	.59	.37	.13	-.07	1.03
15	C	.54	.37	.13	-.09	1.04
16	H	.50	.37	.13	-.12	1.05

None of the steps were significant beyond the .05 level for a two tailed test.

TABLE 80.--MULTIPLE REGRESSION ANALYSIS - MALE INSTRUMENTAL - 16PF - G+R CRITERION

Step	Variable	F Ratio	R	R^2	R'	
1	B	3.27	.22	.05	.05	1.50
2	G	2.19	.25	.06	.05	1.50
3	Q1	1.75	.28	.08	.05	1.50
4	A	1.41	.29	.08	.04	1.50
5	F	1.25	.30	.09	.03	1.51
6	H	1.21	.33	.11	.03	1.51
7	C	1.15	.34	.12	.03	1.51
8	N	1.04	.35	.12	.02	1.52
9	Q3	.95	.36	.13	.01	1.53
10	Q4	.91	.37	.14	.00	1.53
11	G (removed)	1.03	.37	.14	-.01	1.55
12	L	.95	.38	.14	.01	1.53
13	E	.86	.38	.14	-.01	1.54
14	O	.79	.38	.15	-.02	1.55
15	Q2	.73	.39	.15	-.04	1.56
16	M	.67	.39	.15	-.05	1.58
17	I	.62	.39	.15	-.07	1.59

None of the steps were significant beyond the .05 level for a two tailed test.

TABLE 81.--MULTIPLE REGRESSION ANALYSIS - MALE INSTRUMENTAL - 16 PF - G-R CRITERION

Step	Variable	F Ratio	R	R^2	R'	
1	N	2.66	.20	.04	.04	1.24
2	O	2.46	.27	.07	.06	1.23
3	M	2.01	.29	.09	.06	1.23
4	C	1.79	.32	.10	.06	1.22
5	E	1.55	.33	.11	.05	1.23
6	H	1.49	.36	.13	.06	1.23
7	Q4	1.33	.37	.13	.05	1.23
8	G	1.19	.37	.14	.04	1.24
9	B	1.10	.38	.15	.03	1.24
10	Q3	.98	.38	.15	.01	1.25
11	Q1	.88	.38	.15	-.00	1.26
12	A	.80	.39	.15	-.02	1.27
13	L	.73	.39	.15	-.04	1.29
14	I	.67	.39	.15	-.05	1.30
15	F	.61	.39	.15	-.07	1.31

None of the steps were significant beyond the .05 level for a two tailed test.

TABLE 82.--MULTIPLE REGRESSION ANALYSIS - MALE INSTRUMENTAL - MAT VARIABLES - R CRITERION

Variables	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
SW-I	.11	.07	.22	.22	.13
CA-U	-.07	.07	-.13	-.11	.12
HO-U	-.10	.08	-.17	-.18	.13
FR-U	.09	.17	.16	.15	.13
SE-U	-.17	.16	-.15	-.15	.13
AS-U	-.12	.07	-.22	-.22	.13
SW-U	.12	.08	.20	.21	.13
FR-I	.07	.06	.17	.16	.12
SE-I	.06	.07	.11	.10	.13
SS-I	.05	.07	.08	.08	.12

Constant = -.80

$R = .52$ $R^2 = .27$ $R' = .40$ $F = 2.13^*$

$p<.05^*$

CROSS VALIDATION

STUDENT TEACHERS	
Male Instrumental	-.05
Male Choral	-.78
Female Instrumental	.20
Female Choral	.15
MUSIC EDUCATORS	
Male Choral	-.39
Female Instrumental	-.29
Female Choral	.12

Cross validation to the other music educator groups produced negative relationships to the male choral (r= -.39) and female instrumental (r= -.29) and a positive but insignificant correlation to the female choral group (r=.12). Because these were not significant cross validations, the null hypotheses were not rejected. The finding was that the motivational variables which are related to the R criterion for the male instrumental music educator were different from those for other subgroups of music educators indicating that each subgroup was homogeneous, and the total sample was heterogeneous.

Table 83 shows the multiple regression analysis for investigating the relationships between the motivational variables and the G criterion. None of the steps were significant, and the null hypothesis was not rejected. The finding was that the significance of the relationships could not be improved for the male instrumental group. There were no significant relationships between motivational variables and G.

Table 84 shows the multiple regression equation which investigated the relationships between the motivational variables and the G+R criterion. This equation produced a coefficient of R=.41 which was significant beyond the .05 level indicating that 17 percent of the variance in the criterion was accounted for by the motivational variables. The null hypothesis was rejected, and the significance of the relationships was improved by stratifying the sample into subgroups. There were significant relationships between motivational variables and G+R.

Cross validation to the other teacher type groups was insignificant but produced positive coefficients for the female instrumental and female choral groups and a negative coefficient for the male choral group. The cross validation null hypotheses failed to be rejected.

TABLE 83.--MULTIPLE REGRESSION ANALYSIS - MALE INSTRUMENTAL - MAT VARIABLES - G CRITERION

Step	Variable	F Ratio	R	R^2	R'	
1	CA-U	2.33	.18	.03	.03	.98
2	NA-I	2.03	.24	.06	.04	.97
3	SW-U	1.97	.29	.08	.06	.97
4	HO-U	1.66	.31	.10	.05	.97
5	AS-I	1.46	.32	.11	.05	.97
6	SW-I	1.32	.34	.11	.04	.97
7	FR-U	1.18	.35	.12	.03	.98
8	MA-U	1.09	.36	.13	.03	.98
9	AS-U	1.00	.37	.13	.02	.99
10	SE-U	.93	.37	.14	.01	.99
11	SE-I	.86	.38	.14	-.01	1.00
12	PG-I	.79	.38	.15	-.02	1.01
13	CA-I	.73	.39	.15	-.04	1.01
14	NA-U	.68	.39	.15	-.05	1.02
15	FR-I	.62	.39	.15	-.07	1.03
16	PG-U	.58	.39	.15	-.09	1.04
17	SS-I	.54	.39	.16	-.11	1.05
18	MA-I	.50	.39	.16	-.13	1.06
19	HO-I	.47	.39	.16	-.15	1.07

None of the steps were significant beyond the .05 level for a two tailed test.

BEST COPY AVAILABLE

TABLE 84.--MULTIPLE REGRESSION ANALYSIS - MALE INSTRUMENTAL - MAT VARIABLES - G+R CRITERION

Variables	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
CA-U	-.18	.11	-.21	-.20	.12
SE-U	-.10	.09	-.14	-.14	.13
AS-U	-.18	.10	-.22	-.21	.12
SW-U	.18	.11	.20	.20	.13
NA-I	-.11	.10	-.15	-.14	.12

$R = .41$ $R2 = .17$ $R' = .34$ $F = 2.56*$

$p<.05*$

CROSS VALIDATION

STUDENT TEACHER	
none	
MUSIC EDUCATOR	
Male Choral	-.22
Female Instrumental	.10
Female Choral	.12

Table 85 shows the multiple regression equation for investigating the relationships between the motivational variables and the G-R criterion. This equation produced a coefficient of R=.44 which was significant beyond the .05 level indicating that 19 percent of the variance in the criterion was accounted for by the motivational variables alone. The null hypothesis was rejected, and significance was improved for this group: there were significant relationships between motivational variables and G-R, the discrepancy criterion.

Cross validation to the other teacher type groups was insignificant and produced negative coefficients. The finding was that the motivational variables which were related to the G-R criterion for the male instrumental group were not related to success for the other teacher type subgroups. The male instrumental group was different from the other teacher type groups in terms of motivation.

MALE CHORAL AND FEMALE INSTRUMENTAL

When the number of variables approaches the number of subjects in a group, the multiple correlation coefficient must approach R=1.00. This requirement invalidated the Phase Two investigations for the male choral (N=26) and the female instrumental (N=12) groups. The number of variables approached the number of subjects in the male choral group and exceeded the number of subjects in the female instrumental group. Therefore, the findings of the regression analyses for investigating the relationships among personality, motivational, and criterion variables were meaningless data. The multiple R's were spuriously high making useful interpretations of the data impossible.

TABLE 85.--MULTIPLE REGRESSION ANALYSIS - MALE INSTRUMENTAL - MAT VARIABLES - G-R CRITERION

Variables	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
SW-I	-.17	.08	-.25	-.25	.13
HO-U	.20	.09	.26	.27	.13
FR-U	-.13	.09	-.19	-.18	.12
AS-U	.07	.09	.10	.09	.12
FR-I	-.07	.07	-.13	-.12	.12
NA-I	-.08	.08	-.13	.12	.12

Constant = 1.00

$R = .44$ $R^2 = .19$ $R' = .35$ $F = 2.39$*

$p<.05$*

CROSS VALIDATION

MUSIC EDUCATOR

Male Choral	-.20
Female Instrumental	-.52
Female Choral	-.04

FEMALE CHORAL

The female choral group included the largest number of subjects of the teacher type groups (N=103). Table 86 shows the equation for investigating the relationships between the personality variables and the R criterion. This equation produced a multiple R of .28 which was significant beyond the .05 level indicating that 8 percent of the variance in the criterion was accounted for by the personality variables. The null hypothesis that there was no relationship between personality and R was rejected.

Cross validation appeared promising although insignificant. The only negative cross validation r was to the female instrumental music student teacher group while the correlations were positive to all the other groups. The cross validation null hypotheses failed to be rejected.

Table 87 shows the multiple regression equation for investigating the relationships between the personality variables and the G criterion. The equation was significant beyond the .05 level thus rejecting the null hypothesis. The multiple correlation coefficient was respectable (R=.36) which indicated that 13 percent of the variance in $\bar{G}$ was accounted for by the personality variables for the female choral group.

Cross validation to the other music educator groups was not significant, and the null hypotheses failed to be rejected. With more subjects, cross validation appeared promising.

TABLE 86.--MULTIPLE REGRESSION ANALYSIS - FEMALE CHORAL - 16PF - R CRITERION

Variables	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
G	.05	.04	.12	.15	.13
O	-.05	.03	-.17	-.19	.11
Q3	.03	.05	.07	.09	.13

Constant = -.54

$R = .28$ $R^2 = .08$ $R' = .25$ $F = 2.86^*$

$p<.05^*$

CROSS VALIDATION

STUDENT TEACHER	
Female Choral	.12
Female Instrumental	-.30
Male Choral	.04
Male Instrumental	.01
MUSIC EDUCATOR	
Male Instrumental	.13
Male Choral	.07
Female Instrumental	.69

TABLE 87.--MULTIPLE REGRESSION ANALYSIS - FEMALE CHORAL - 16PF - G CRITERION

Variables	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
B	.05	.05	.07	.07	.10
E	.03	.03	.12	.13	.10
G	.07	.04	.16	.21	.13
O	-.08	.03	-.24	-.27	.16
Q2	-.06	.13	-.17	-.17	.10
Q3	-.07	.15	-.16	-.21	.13

Constant = .44

$R = .36$ $R^2 = .13$ $R' = 29$ $F = 2.33$*

$p < .05$*

CROSS VALIDATION

MUSIC EDUCATOR	
Male Choral	.14
Male Instrumental	-.02
Female Instrumental	-.15

Table 88 shows the multiple regression equation for investigating the relationships between personality variables and the G+R criterion for the female choral group. This equation was significant and produced a multiple R of .39 indicating that 15.21 percent of the variance in the criterion was accounted for by the equation. The null hypothesis was rejected. Personality variables were related to G+R for the female choral group.

Cross validation appeared extremely promising although insignificant, and the null hypotheses were not rejected. The finding was that the significance of the relationships was improved by stratifying the sample for the female choral group.

Table 89 shows the multiple regression equation which best investigated the relationships between personality variables and the G-R criterion. Only Factor Q3 entered the equation (Self Sentiment Integration), and that while the equation was significant, one does not usually consider a single variable multiple regression equation as having much usefulness. Therefore, the null hypothesis was rejected beyond the .05 level. Cross validation coefficients were negative and insignificant. The finding was that the significance of the relationships to the G-R criterion for the female choral group did improve by stratifying the groups but this improvement was minimal. Personality variables were related to G-R for the female choral group. No significant equations could be derived for predicting G-R from personality variables for the total derivation sample.

TABLE 88.--MULTIPLE REGRESSION ANALYSIS - FEMALE CHORAL - 16PF - G+R CRITERION

Variables	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
B	.14	.11	.12	.13	.10
F	.04	.04	.11	.12	.12
G	.10	.15	.19	.21	.11
I	.04	.07	.06	.07	.11
M	-.06	.06	-.10	-.12	.12
N	-.06	.06	-.09	-.10	.11
O	-.13	.04	-.31	-.34	.11
Q2	-.03	.05	-.06	-.07	.11

Constant = -.63

$R = .39$ $R^2 = .15$ $R' = .30$ $F = 2.11^*$

$p<.05^*$

CROSS VALIDATION

MUSIC EDUCATOR	
Male Instrumental	.15
Male Choral	.16
Female Instrumental	.42

TABLE 89.--MULTIPLE REGRESSION ANALYSIS - FEMALE CHORAL - 16PF - G-R CRITERION

Variables	b Coefficient	SE of b	Partial R	β Coefficient	SE of β
Q3	-.10	.05	-.20	-.20	.10

Constant = 1.25

R = .20 R^2 = .04 R' = .20 F = 4.21*

p<.05*

	CROSS VALIDATION
MUSIC EDUCATOR	
Male Instrumental	-.04
Male Choral	-.28
Female Instrumental	-.12

Tables 90 through 93 show the multiple regression equations for investigating the relationships between the motivational variables and the R, G, G+R, and G-R criteria. All of the equations were significant beyond the .05 level providing respectable multiple R's. For the R criterion, the multiple R was .46 which indicated that 21.42 percent of the variance in the criterion was accounted for by the motivational variables. The G criterion's equation produced a multiple R of .50 which indicated that 24.82 percent of the variance in G was accounted for by the motivation variables. The G+R criterion's equation produced a multiple R of .48 which accounted for 23.38 percent of the variance, and the G-R criterion's equation produced a multiple R of .48 accounting for 22.57 percent of the variance. The null hypotheses were rejected indicating that relationships existed among the motivational and criterion variables. The finding was that the significance of the relationships was improved by stratifying the female choral group from the sample and developing separate multiple regression equations for investigating the relationships among motivational and criterion variables.

Cross validation equations were not significant, and the null hypotheses were not rejected. Additional subjects would be needed for more meaningful cross validations.

Summary

The hypotheses that were investigated by partitioning the sample into teacher type groups were that there were no significant relationships between personality and motivational variables and the 4 criteria of music teaching success. For any significant multiple regression equations, it was further hypothesized that these relationships would not cross validate to the music student teacher type subgroups or to the other music educator type groups. Furthermore, it was of concern to the study to determine whether the significance of the relationships would be improved by partitioning the sample into teacher type subgroups

TABLE 90.--MULTIPLE REGRESSION ANALYSIS - FEMALE CHORAL - MAT VARIABLES - R CRITERION N = 103

Variables	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
SW-I	.11	.06	.20	.20	.10
CA-U	.12	.06	.19	.19	.10
HO-U	-.05	.06	-.09	-.08	.10
SE-U	-.14	.07	-.21	-.22	.11
SS-U	.08	.07	.11	.11	.10
PG-U	.03	.06	.06	.06	.11
AS-U	-.12	.06	-.20	-.20	.10
SW-U	-.09	.06	-.17	-.18	.11
CA-I	-.12	.06	-.21	-.21	.10
FR-I	-.04	.16	-.08	-.07	.10
NA-I	.14	.06	.27	.26	.10
SS-I	-.05	.07	-.08	-.08	.10
AS-I	-.07	.06	-.13	-.13	.10

Constant = 1.05

R = .46 R^2 = .21 R' = .33 F = 1.87*

p<.05*

BEST COPY AVAILABLE

CROSS VALIDATION

STUDENT TEACHER	
Female Choral	-.17
Female Instrumental	.03
Male Choral	-.94
Male Instrumental	-.07
MUSIC EDUCATOR	
Male Instrumental	.20
Male Choral	-.21
Female Instrumental	-.07

TABLE 91.--MULTIPLE REGRESSION ANALYSIS - FEMALE CHORAL - MAT VARIABLES - G CRITERION N = 103

Variables	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
SW-I	.08	.06	.14	.14	.11
CA-U	-.06	.17	-.10	-.10	.11
HO-U	-.11	.16	-.19	-.18	.10
FR-U	.09	.06	.17	.17	.10
NA-U	-.01	.05	-.03	-.03	.11
SS-U	.14	.07	.21	.21	.10
MA-U	.06	.06	.12	.12	.11
PG-U	-.03	.06	-.05	-.05	.11
SW-U	.08	.06	.15	.16	.11
CA-I	.02	.06	.04	.04	.10
HO-I	-.03	.06	-.05	-.05	.11
SE-I	.02	.06	.03	.03	.10
SS-I	.12	.07	.17	.18	.11
MA-I	-.14	.08	-.19	-.22	.12
PG-I	-.04	.05	-.10	-.09	.10
AS-I	-.05	.06	-.10	-.09	.10

Constant = -.93

R = .50 R^2 = .25 R' = .34 F = 1.77*

p<.05*

CROSS VALIDATION

MUSIC EDUCATOR	
Male Choral	-.37
Male Instrumental	-.08
Female Instrumental	.04

TABLE 92.--MULTIPLE REGRESSION ANALYSIS - FEMALE CHORAL - MAT VARIABLES - G+R CRITERION N = 103

Variables	b Coefficients	SE of b	Partial R	β Coefficients	SE of β
SW-I	.19	.09	.24	.23	.10
CA-U	.06	.09	.07	.07	.10
HO-U	-.16	.08	-.20	-.19	.10
FR-U	.08	.07	.12	.11	.10
SE-U	-.13	.09	-.15	-.14	.10
SS-U	.22	.10	.23	.22	.10
MA-U	.08	.08	.11	.11	.10
AS-U	-.10	.19	-.12	-.12	.10
CA-I	-.10	.08	-.12	-.12	.10
NA-I	.13	.08	.16	.16	.10
SS-I	.06	.11	.06	.06	.11
MA-I	-.12	.10	-.13	-.13	.11
PG-I	-.07	.07	-.10	-.10	.11
AS-I	-.12	.08	-.16	-.16	.10

Constant = -.32

R = .48 R^2 = .23 R' = .35 F = 1.92*

p<.05*

CROSS VALIDATION

MUSIC EDUCATOR	
Male Choral	-.36
Male Instrumental	.08
Female Instrumental	-.18

TABLE 93.--MULTIPLE REGRESSION ANALYSIS - FEMALE CHORAL - MAT VARIABLES - G-R CRITERION N = 103

Variables	b Coefficients	SE of b	Partial N	β Coefficients	SE of β
CA-U	-.20	.09	-.24	-.23	.10
HO-U	-.04	.08	-.06	-.05	.10
FR-U	.11	.08	.16	.15	.10
SE-U	.16	.10	.17	.18	.11
SS-U	.07	.10	.08	.07	.10
MA-U	.05	.08	.07	.07	.10
PG-U	-.06	.08	-.08	-.08	.11
AS-U	.13	.09	.16	.16	.10
SW-U	.18	.08	.22	.26	.12
CA-I	.14	.08	.19	.18	.10
HO-I	-.04	.09	-.05	-.05	.11
NA-I	-.13	.18	-.17	-.17	.10
SS-I	.20	.10	.21	.21	.11
MA-I	-.17	.10	-.17	-.19	.12

Constant = -2.39

$R = .48$ $R^2 = .23$ $R' = .34$ $F = 1.83^*$

$p<.05^*$

CROSS VALIDATION

MUSIC EDUCATOR	
Male Choral	-.11
Male Instrumental	.06
Female Instrumental	.22

as indicated by the size of the multiple correlation coefficients.

Significant relationships existed between personality and the R criterion for the male instrumental and female choral groups; between personality and the G criterion for the female choral group; between personality and the G+R criterion for the female choral group; and between personality and the G-R criterion for the female choral group. Significant relationships existed between motivation and the R criterion for the male instrumental and female choral groups; between motivation and the G criterion for the female choral group; between motivation and the G+R criterion for the male instrumental and female choral groups; and between motivation and the G-R criterion for the male instrumental and female choral groups.

None of the above significant relationships cross validated to either the music student teacher type subgroups or to the music education teacher type subgroups. Interestingly, however, was that a hint of cross validation to the music student teacher female choral group was evident in the findings for the personality domain. There was evidence that with larger numbers of subjects and more valid criteria the significance and power of the relationships would be improved.

The attempt to cross validate the music educator teacher type subgroup findings with the other subgroups was to further investigate whether personality and motivational variables which were significantly related to success for one subgroup were related to the same criterion for other subgroups. The fact that none of these cross validation r's was significant indicated that the music educator sample was heterogeneous as far as the personality and motivation. In other words, each subgroup of music teacher type was homogeneous requiring different combinations of the personality and motivational variables which contributed to the multiple regression equations, e.g. the significance of the relationships. The finding was that teacher type subgroups of music educators were different in personality and motivation as these variables related to music teaching success criteria.

In general, the significance of the relationships was strengthened by partitioning the sample into music teacher type subgroups although these could not be cross validated on like samples as they were for the derivation sample. The significant relationships between the personality variables and the music teaching success criteria were not substantiated on the cross validation sample as in Phase One.

The multiple R's for investigating the relationships between the motivational variables and the G and G-R criteria for the derivation sample were .35 and .34 respectively with cross validations of .40 and .28. No significant relationships between the personality variables and the R or G+R criteria appeared to exist. Partitioning of the sample improved the significance of the relationships between the motivational variables and G-R: male instrumental .44 and female choral .48. These relationships also improved to the G criterion for the female choral group .50 but not for the male instrumental group. While the R and G+R criteria were not related to motivation for the derivation sample, the significance of the relationships were improved for the two teacher type subgroups (male instrumental and female choral): their equations were all significant beyond the .05 level. The finding was that, in all but one case, partitioning of the sample into teacher type subgroups for investigating the relationships between the motivational variables and music teaching success criteria helped to further identify the significant relationships.

PERSONALITY AND MOTIVATION PROFILES

Personality and motivational variables as measured by the instruments used in the study refer to major categories of traits, interests, feelings, and drives. These may or may not be interrelated depending upon the success criteria used to investigate the relationships. In investigating the relationships among personality and motivational variables and music teaching success criteria, the variables which contribute to the regression equations are weighted, and these weights are applied to the cross validation subjects' scores on the personality and motivational variables as was done in Phase One analyses. Inferences were then drawn about the significance of the relationships among the experimental and criterion variables.

This section is an attempt to highlight the possible profile differences between successful and unsuccessful male and female groups of music educators. Profiles were inferred from the following multivariate analyses rather than from analyses which treated each experimental variables as a discrete entity. The common technique used in previous studies in music education in order to study profile differences between groups investigated the differences between mean scores on the variables which were significant at or beyond the .05 or .01 levels using chi square or t-test techniques. These approaches appear to be the natural way to describe such differences between two groups and seem to present an intuitively meaningful and sensible portrayal of the group profile differences. The technical difficulty with this approach is that the danger of obtaining a distorted picture of the group differences tends to increase as the correlations among the variables become larger. Tatsuoka discusses the difficulty of using mean differences for profile comparisons.

> . . . as the number of variables increase, the difficulty of interpreting differences between two or more groups on each variable taken singly will become more and more serious . . . Besides the question of interpretability, there is the more technical matter of the level of significance of the differences becoming "muddied up"

> when t-tests are done on many correlated variables . . . But even the difficulty of interpretation alone should offer sufficient grounds for looking askance at the approach of examining the variables one at a time . . .
>
> We must therefore look for an alternative way to describe group differences. One such alternative is to construct a linear combination . . . of the set of variables that will maximally differentiate among the groups in question. We then see how the groups are ranked in terms of this linear combination (essentially a new, transformed variable), and also examine the relative weights assigned to the different variables in forming the linear combination. These two observations, taken together, enable us to describe the nature of group differences . . . The pattern of weights here indicates how much (or little) each factor contributes, and in which direction, to the differentiation between our groups (Tatsuoka, 1970:3).

Male and female music educators were stratified into successful and unsuccessful gain (G) and rating (R) criterion groups. The upper and lower 25 percent of the samples respectively were used for both sex groups.

Two group stepwise discriminant analyses were conducted between the successful and unsuccessful male and female groups for both criteria yielding 8 equations or sets of b weights which highlighted profile differences separately for the personality and motivational variables. These weights are analoguous to the partial regression weights in a multiple regression equation since each weight represents the relative importance of that variable with the effects of the other variables partialled out. By using the stepwise technique, only those variables which contributed to the significance of the linear combination entered or were removed from the equation.

The null hypotheses tested were that there would be no significant profile differences between successful and unsuccessful groups of music educators for the G and R criteria when investigated separately by sex. The null hypotheses were rejected beyond the .05 level for 7 of the 8 profile analyses. The null hypothesis could not be rejected for the motivational profile differences between successful and unsuccessful male music educators on the G criterion.

Tables 94 and 95 show the weights for the successful and unsuccessful personality and motivational profile group comparisons for the following: (1) Male Gain, (2) Female Gain, (3) Male Rating, and (4) Female Rating. The coding for the successful and unsuccessful groups were: successful = 2.00, unsuccessful = 1.00; hence, variables on which successful music educators had the higher means tended to receive positive weights, while those on which the unsuccessful music educators had the higher means tended to receive negative weights in the equations. Profile comparisons are discussed within the equations as well as between equations.

The weights for Equation 1, Table 94, indicate that the successful male music educator in G is more intelligent (B), stubborn (E), conscientious (G), astute and shrewd (N), experimenting (Q1), and socially precise (Q3) than his unsuccessful colleague in G who is calmer (C). The most important profile differences appear to be in intelligence, stubbornness, shrewdness, and social precision which characterized the successful male group.

For the motivation variables, no equation was significant beyond the .05 level for the male music educators on the G criterion. It appears that the profile differences between successful and unsuccessful male music educators in G were not of sufficient magnitude to identify motivation profiles for these two groups.

The weights in Equations 2, Tables 94 and 95 indicate that the successful female music educator in G tended to be stubborn (E), conscientious (G), somewhat apprehensive (O), experimenting (Q1), more self sufficient and less group oriented (Q2), is interested in security and avoiding financial loss (FR-U), is interested in social approval and concerned about her general reputation (SS-U, SS-I), had a higher interest in the opposite sex (SW-U, SW-I), is more assertive and driven towards status (AS-U, AS-I), is more interested in her career and had a desire to excel in competition (CA-I), and is more interested in getting things accomplished (SE-I) than her unsuccessful colleague in G who is more astute and

TABLE 94.--COMPARISONS OF SUCCESSFUL AND UNSUCCESSFUL MUSIC EDUCATOR PROFILES FOR R AND G CRITERIA SEPARATELY BY SEX FOR THE 16PF TEST - b WEIGHTS

Equations	A	B	C	E	F	G	H	I	L	M	N	O	Q1	Q2	Q3	Q4
1. MG		.04	-.03	.04		.02					.06		.03		.06	
2. FG	-.01		-.01	.04		.05					-.07	-.03	-.03	-.05		
3. MR		.07			-.03		.02		-.05	.04						
4. FR					.02	.04			.04			-.07	-.07	.03	.02	

1. = Male Gain R = .54 R^2 = .29 R' = .42 F = 2.21*
2. = Female Gain R = .52 R^2 = .27 R' = .41 F = 2.30* Males N = 46
3. = Male Rating R = .50 R^2 = .25 R' = .42 F = 2.63* Females N = 58
4. = Female Rating R = .45 R^2 = .20 R' = .40 F = 2.33*

$p<.05$*

TABLE 95.--COMPARISONS OF SUCCESSFUL AND UNSUCCESSFUL MUSIC EDUCATOR PROFILES FOR R AND G CRITERIA SEPARATELY BY SEX FOR THE MAT TEST - b WEIGHTS

	Unintegrated										Integrated									
Equations	CA	HO	FR	NA	SE	SS	MA	PG	AS	SW	CA	HO	FR	NA	SE	SS	MA	PG	AS	SW
1. MG																				
2. FG	-.05	-.05	.08	-.02	-.03	.11			.04	.07	.04		.02	.01	.05	.05	-.06		.04	.04
3. MR	-.06	-.10	-.11	-.04	-.10	-.06	.03		-.11		-.06	.04		-.06	.05				-.07	.06
4. FR	.06				-.09	.04		-.04	-.05		-.05			.09		-.03	-.04			.05

1. = Male Gain not significant
2. = Female Gain R = .65 R^2 = .43 R' = .47 F = 1.92* Males N = 46
3. = Male Rating R = .70 R^2 = .49 R' = .53 F = 2.09* Females N = 58
4. = Female Rating R = .57 R^2 = .32 R' = .44 F = 2.25*

$p<.05$*

shrewd (N), had more unconscious interests in her career (CA-U) and home (HO-U), and is attracted to the opposite sex (MA-I). The unsuccessful female music educator in G appeared to be more interested in being a wife and mother, perhaps, than in succeeding in a music teaching career. The successful female music educator's profile appears to be somewhat similar to that of the liberated woman, at least as far as the gain criterion was concerned.

Equations 1 and 2, Table 94, show some moderating sex differences. It appears that the successful male music educator in G is shrewd and astute (N) and somewhat experimenting (Q1) as is the unsuccessful female music educator in G. On the other hand, the successful male and female music educator in G is stubborn (E) and conscientious (G) while the unsuccessful male and female in G is somewhat calm (C).

For the R criterion, the weights from Equations 3, Tables 94 and 95 indicate that the successfully rated male music educator is more intelligent (B), somewhat more venturesome (H), more imaginative (M), is interested in the opposite sex and beauty (MA-U, SW-I), is more attached to his home (HO-I), and is interested in getting things accomplished (SE-I) where his unsuccessfully rated colleague is somewhat enthusiastic (F), more suspicious (L), is interested in but appeared to have a conflict with his career (CA-U, CA-I) as related to his unconscious attachments to his home (HO-U) as well as his unconscious interests in security (FR-U), social approval, and concerns about his general reputation (SS-U). Furthermore, the unsuccessfully rated male music educator appeared to experience some conflicts in his interests in sensual indulgences (NA-U, NA-I) and drives toward status (AS-U, AS-I) as indicated by the Unintegrated and Integrated conscious motive strengths.

The weights for the unsuccessfully rated male group were considerably higher than those for the successfully rated male group substantiating to some extent the suspicious trait found in Factor L. The successfully rated male music educator, while interested in and receiving satisfaction from his home and his wife, is interested in getting things accomplished which he appeared to do in more imaginative and venturesome ways.

The successfully rated female music educator is somewhat more enthusiastic (F), conscientious(G), suspicious (L), self sufficient (Q2), socially precise (Q3), is unconsciously interested in her career (CA-U), concerned about her general reputation and interested in social approval (SS-U), is interested in sensual indulgences (NA-I), and the opposite sex (SW-I) while her unsuccessfully rated colleague is more apprehensive (O), experimenting (Q1), has more unconscious interests in getting things accomplished (SE-U), drives toward status (AS-U), and tends to be more defensive and hostile (PG-U) while being more consciously interested in her career (CA-I), in social approval (SS-I), and in the opposite sex (MA-I). The overall profile for the successfully rated female indicates that she is concerned about the social niceties while being conscientious and enthusiastic about her career. On the other hand, the unsuccessfully rated female music educator appeared to be more defensive and apprehensive as well as concerned about status which may have an effect on her ability to obtain high ratings.

By reading Equations 3 and 4 for Tables 94 and 95, interactions or moderators can be noted. The successfully rated female and unsuccessfully rated male music educator is somewhat enthusiastic (F), more suspicious (L), unconsciously interested in teaching music (CA-U) as well as social approval and reputation (SS-U), and is more consciously interested in and received satisfactions from sensual indulgences of all kinds (NA-I). In both these tables, the successfully rated male and female is similar only in attachments to the opposite sex (SW-I) while the unsuccessfully rated male and female tends to have unconscious interests in getting things accomplished (SE-U) and assertiveness or status (AS-U) while being somewhat interested in career (CA-I).

Comparing across criteria (Equations 1 and 3 in Tables 94 and 95), the successful male music educator is intelligent (B). No similarities were noted for the unsuccessful male music educator. The successful female music educator, (Equations 2 and 4, Tables 94 and 95), is conscientious (G), interested in social approval and concerned about reputation (SS-U), somewhat interested in sensual indulgences (NA-I), and interested in the opposite sex (SW-I). The unsuccessful female music educator is apprehensive (O), self sufficient (Q1), unconsciously interested in getting things accomplished (SE-U), and interested in the opposite sex (MA-I). It appears that the successful and unsuccessful profiles for the G and R criteria differ as far as the motivational and personality variables are concerned.

Sex Success Profiles

Additional analyses were undertaken to investigate the profile differences between male and female music educators who were classified similarly on the G and R criteria. In these analyses, two group stepwise discriminant analyses were undertaken between the male and female groups for: (1) successful gain, (2) unsuccessful gain, (3) successful rating, and (4) unsuccessful rating for both the personality and motivational variables. The null hypotheses tested were that there would be no significant profile differences between the male and female groups as stratified by success. All 8 equations were significant beyond the .05 level, and the null hypotheses were rejected. Tables 96 and 97 show the unstandardized weights for the 8 equations. The coding for sex was male = 1.00, female = 2.00; hence, variables on which females had the higher means tended to receive positive weights, while those on which males had the higher means tended to receive negative weights in the equations.

By examining the weights in Equations 1 for Tables 96 and 97, it appears the successful male music educator in G is intelligent (B), conscientious (G),

TABLE 96.--COMPARISONS OF MALE AND FEMALE MUSIC EDUCATOR PROFILES FOR LEVEL OF SUCCESS SEPARATELY BY R AND G CRITERIA - UNSTANDARDIZED REGRESSION WEIGHTS - 16PF TESTS

Equations	A	B	C	E	F	G	H	I	L	M	N	O	Q1	Q2	Q3	Q4
1. High G		-.02		-.01	.04	-.03	-.02	.07	-.03	-.03	-.03	-.03	-.03		-.04	.03
2. Low G	.01	.06	.00	-.01	-.01	-.02	.00	.03	-.08	-.03	.09	.02	.05	.01	-.02	.02
3. High R	-.02	-.07	.01	-.04				.07	.02	-.02	.04	-.02	.02	-.01	-.06	.02
4. Low R	-.01	.06	-.01	.01	-.02	-.06	.01	.07	-.02	-.01		.01	-.01	-.02		.03

1. High G R = .65 R^2 = .42 R' = .49 F = 2.12*
2. Low G R = .72 R^2 = .51 R' = .56 F = 2.32*
3. High R R = .64 R^2 = .41 R' = .48 F = 2.06*
4. Low R R = .67 R^2 = .45 R' = .51 F = 2.14*

Females N = 58
Males N = 46

p<.05*

TABLE 97.--COMPARISONS OF MALE AND FEMALE MUSIC EDUCATOR PROFILES FOR LEVEL OF SUCCESS SEPARATELY BY R AND G CRITERIA - UNSTANDARDIZED REGRESSION WEIGHTS - MAT TESTS

	Unintegrated										Integrated									
Equations	CA	HO	FR	NA	SE	SS	MA	PG	AS	SW	CA	HO	FR	NA	SE	SS	MA	PG	AS	SW
1. High G		-.05	.10			.13		-.05	.04	.10	-.01	.05			-.04	-.04	-.04	-.05	.03	.02
2. Low G	-.05	.03	-.09			-.09		-.08	-.02	.05	-.08			-.04	-.08	-.10		-.03	-.03	
3. High R	.03	.08		.07		.08		-.05	.07	.05	-.03		-.02	.08	-.06	-.05				.02
4. Low R	-.09	-.03	-.05	.07		.08		.06		.07		.07	.03					-.07	.04	

1. High G R = .66 R^2 = .43 R' = .48 F = 1.99*
2. Low G R = .65 R^2 = .42 R' = .49 F = 2.10*
3. High R R = .65 R^2 = .43 R' = .50 F = 2.19*
4. Low R R = .61 R^2 = .38 R' = .47 F = 2.19*

Females N = 58
Males N = 46

p<.05*

suspicious (L), imaginative (M), shrewd and astute (N), apprehensive (O), experimenting (Q1), socially precise (Q3), interested in his home (HO-U), tends to be defensive and hostile (PG-U, PG-I), interested in accomplishing things (SE-I), interested in social approval and concerned about his reputation (SS-I), and is interested in the opposite sex (MA-I). On the other hand, the successful female music educator in G is enthusiastic (F), sensitive (I), tense (Q4), has unconscious interests in security (FR-U), social approval and concern about her reputation (SS-U) as well as interests in the opposite sex (SW-U), and is interested in and receives satisfaction from her home (HO-I) and is assertive and drives herself towards status (AS-I).

The unsuccessful male music educator in G (Equations 2, Tables 96 and 97), is suspicious (L), interested in self indulgences (NA-I) and in getting things accomplished (SE-I) but has some conflicts in his motivations as indicated by the Unconscious Motive Strengths indicating that he is interested in his career (CA-U, CA-I), social approval and reputation (SS-U, SS-I), and is defensive and hostile (PG-U, PG-I). This profile suggests a music educator who is interested in the right things and has certain drives towards these things but is unable to go about achieving these things so that conflicts appear evident. The unsuccessful female music educator in G appears to be intelligent (B), shrewd and astute (N), experimenting (Q1), and unconsciously interested in the opposite sex (SW-U).

By examining the weights for Equations 1 and 2 in Tables 96 and 97, males appear to be similar. The male profiles for both G groups indicates that the male is stubborn (E), conscientious (G), suspicious (L), imaginative (M), socially precise (Q3), defensive and hostile (PG-U, PG-I), interested in his career (CA-I), interested in getting things accomplished (SE-I), and interested in social approval (SS-I). On the other hand, the female profiles indicate that the female is senstivie (I), tense (Q4), and unconsciously interested in the opposite sex (SW-U). Fewer variables were included in the female profile.

By examining the weights in Equations 3, Tables 96 and 97, the profiles for the successfully rated male and female groups emerge. The male is intelligent (B), stubborn (E), socially precise (Q3), unconsciously defensive and hostile (PG-U), interested in accomplishing things (SE-I), and interested in social approval and concerned about his reputation (SS-I). The successful female is sensitive, (I), shrewd and astute (N), unconsciously interested in her home (HO-U) as well as in social approval and reputation (SS-U), assertiveness and status (AS-U), and the opposite sex (SW-U), while having both conscious and unconscious interests in self indulgences (NA-U, NA-I).

The unsuccessfully rated male, Equations 4, Tables 96 and 97, is conscientious (G), unconsciously interested in his career (CA-U) and security (FR-U) while being consciously defensive and hostile (PG-I). The unsuccessfully rated female is intelligent (B), sensitive (I), tense (Q4), unconsciously interested in sensual indulgences (NA-U), social approval and reputation (SS-U), hostility and defensiveness (PG-U), and the opposite sex (SW-U), while being consciously interested in her home (HO-I). The unsuccessfully rated male and female had more unconscious interests and desires than their successfully rated colleagues.

By examining the weights in Equations 1 and 3, Tables 96 and 97, similarities in profiles for criteria can be inferred. The successful male is intelligent (B), stubborn (E), imaginative (M), apprehensive (O), socially precise (Q3), unconsciously defensive and hostile (PG-U), interested in his career (CA-I) as well as being interested in accomplishing things (SE-I) and in social approval and concern about his general reputation (SS-I). The successful female is more sensitive (I), tense (Q4), unconsciously interested in social approval and reputation (SS-U) and status (AS-U) while being interested in and receiving satisfaction from the opposite sex (SW-U, SW-I).

The weights in Equations 2 and 4, Tables 96 and 97 suggest that the unsuccessful male is somewhat less enthusiastic (F), conscientious (G), highly suspicious (L), imaginative (M), unconsciously interested in his career (CA-U), has unsatisfied drives toward security (FR-U), and is generally defensive and hostile (PG-I). The unsuccessful female is intelligent (B), sensitive (I), somewhat apprehensive (O), tense (Q4), and has some unsatisfied needs relating to the opposite sex (SW-U).

The above discussion highlighted the major profile differences in personality traits and motivations between male and female music educators. While the equations were significant beyond the .05 level, the magnitude of the weights was not large indicating that profile differences were in degree.

ADVISING AND COUNSELING MUSIC EDUCATION MAJORS

The fourth subproblem asked the question: what are the implications of the use of personality measures for counseling and advising undergraduate music education students? The answer to this question was one of interpreting the findings.

The first consideration was the validity of the music student teacher rating data. Because more missing ratings were predicted than those that were gathered, a serious validity problem became evident. This negated the use of the music student teacher sample for cross validation purposes in Phase One of the regression analyses as well as any separate investigations between success groups of music student teachers on the personality and motivational variables. With larger numbers of subjects in the student teaching group, cross validation appeared promising as indicated by Phase Two findings which attempted to cross validate the music educator equations for teacher type groups internally as well as externally to like music student teacher groups. While these were insignificant statistically, the hint of cross validation was evident.

The second consideration was the size or magnitude of the multiple regression equations found significant in Phase One. While these equations were statistically significant and well cross validated beyond the .05 level indicating that relationships existed between the motivational variables and the G and G-R criteria, many persons would consider the multiple R's to be too low for making any practical decisions.

> The older literature in testing placed little value on tests with moderate validity coefficients . . . According to this coefficient, validity had to reach .86 before a test was "50 percent better than chance." Tests with validity below .50 were thought to have negligible practical value . . . Coefficients as low as .30 are of definite practical value (Cronbach 1965:349).

For the G criterion, the personality test variables of the 16PF test produced a multiple correlation coefficient of R=.26 which was significant beyond the .05 level but was not statistically significant on cross validation. Even so, this equation accounted for 7 percent of the variance in the criterion. The motivation equation produced a multiple R of .35 and a cross validation r of .40 which accounted for 12.25 percent and 16 percent of the variance in the criterion respectively. Almost one-sixth of a music educator's gain success was significantly related to motivational variables alone. No variables representing musicianship, college training, years of teaching experience, and so forth contributed to this equation.

For the R criterion, the personality trait variables produced a multiple R of .42 which was significant beyond the .05 level and which accounted for 17.64 percent of the variance in the criterion. This equation did not cross validate significantly, but the suggestion of the relationships among personality trait variables and music teaching success was evident. These relationships were further highlighted in the profile section. No significant relationships appeared to exist between the linear combination of the motivational variables and the R criterion.

Relationships existed between the G+R criterion and the personality trait variables which produced a multiple R of .39 which was significant beyond the .05 level and accounted for 15.21 percent of the variance in the criterion. The equation was not significant on cross validation. No significant relationships appeared to exist between the G+R criterion and the motivational variables.

The G-R discrepancy criterion was not related to the personality trait variables since no significant equations could be formulated. But, the motivational variables produced a multiple correlation coefficient of R=.34 which was significant beyond the .05 level and a cross validation coefficient of r=.28. These figures indicated that 11.56 percent and 7.84 percent of the variances were accounted for by the equations respectively.

The above equations indicated that personality trait variables were related to G and R although additional cross validation research will be needed; that motivational dynamic variables were related to G and G-R; and that these equations were somewhat more stable than the personality trait equations because cross validation was significant. When the sample was partitioned into teacher type and sex groups, the relationships were improved although cross validation was impossible. In other words, sex interactions tended to lower the multiple R's for the total sample because of sample heterogeneity. By partitioning the sample into homogeneous teacher type-sex groups, the size of the multiple correlation coefficients was larger. Combine this information with the profile analyses which also generated larger multiple R's, and evidence exists that personality and motivational variables are related to music teaching success. The profile analyses included multiple correlation coefficients ranging from R=.45 to R=.72, which indicated that between 20 and 52 percent of the variances in the profiles could be attributed to personality and motivational variables. All of this was useful information.

Before one can advise and counsel future music educators, he must first be able to describe in some meaningful way the personality and motivational dimensions related to successful music teaching. If these can be defined to some extent, they can be included in advising and counseling potential music educators.

> A test may appreciably improve predictive efficiency if it shows any significant correlation with the criterion, however low. Under certain circumstances, even validities as low as .20 or .30 may justify inclusion of the test in a selection program. For many testing purposes, evaluation of tests in terms of the error of estimate is unrealistically stringent. Consideration must be given to other ways of evaluating the contribution of a test, which take into account the types of decisions to be made from the scores (Anastasi 1968:131).

Predictive efficiency refers to those relationships among the experimental and criterion variables which indicated that significant relationships existed. Therefore, the resulting multiple R's were viewed as validity coefficients. Would better judgments about potential music educators be made by using the information derived from the multiple regression and profile analyses than by not using this information?

A Counseling and Advising Model

The findings of the study indicated that personality and motivational variables were related to music teaching success criteria and that differences exist for teacher type subgroups as well as for the sex and criterion profile groups. An advising model must attempt to generalize both similarities and differences as these relate to music teaching success.

In general, the music educator who is successful in both gain and rating is intelligent although some evidence exists that the unsuccessful female music educator is more intelligent than her successful colleague, is interested in social approval and concerned about reputation, is assertive and driven towards status, is interested in the opposite sex, enjoys sensual indulgences of all kinds, and is somewhat hostile and defensive. The music educator who produced

BEST COPY AVAILABLE

measurable learning in his or her pupils is also more conscientious. On the other hand, the music educator who is highly rated, in addition to the previously mentioned traits and motivations, is interested in security and receives satisfaction from music teaching as a career. The successful male music educator in both gain and rating is also more imaginative and somewhat apprehensive while the successful female is more practical and self assured.

The successful male music educator in both gain and rating in addition to the above similarities tends to be stubborn, imaginative, apprehensive, socially precise, interested in his home, receives satisfaction from music teaching as a career, is interested in his accomplishments, social approval, and status. He appears to be somewhat extraverted. If he desires to produce more pupil measurable learning, he is more shrewd and astute, experimenting, and interested in the opposite sex. On the other hand, if he is more interested in being rated successfully by his pupils, peers, or superiors, he tends to be more participating, enthusiastic, venturesome, concerned about status and more ambitious. He appears to be more traditional in his interests, drives, and behaviors.

In addition to the previously mentioned behaviors and interests for both males and females who were successful in both gain and rating, the successful G+R female appears to be more enthusiastic, conscientious, sensitive, practical, forthright, self assured, group oriented, tense, and is somewhat interested in her home but experiences conflicts between her unconscious and conscious interests in music teaching as a career as well as her interests in the opposite sex. These conflicts could be interpreted to be related to what she considers to be her duties at home versus her obligations to the job.

The female music educator who is successful in obtaining measurable pupil learning, in addition to the above traits and interests, is more stubborn and interested in accomplishing things. On the other hand, the female music educator who is rated successfully appears to be more suspicious, conscientious, shrewd and astute, experimenting, and socially precise. She is more positive in developing the interpersonal relationships required for obtaining positive rating evaluations.

The overall attributes leading to music teaching success appear to be those that can be identified during the undergraduate music teacher training period for prospective teachers. While it is evident that advising and counseling procedures would need to consider the major differences in personality traits and motivations between the sexes as these relate to the criteria used in the study, some generalizations can be offered.

The interpersonal relationship variables seem to be most important. Can the potential music educator relate to elementary and secondary pupils? Is he creative in structuring music lessons for children? Does he experiment with different approaches or ideas? Is he concerned about his appearance and how others think about him? Does he like to work with others to solve mutual problems? Is he group oriented or does he prefer to work alone? Does he have a happy self concept in that he is satisfied with himself and looks forward to the future? Is he conscientious?

While stubbornness may be a quality for producing pupil music achievement, it may not get the music education major graduated nor his first job. Therefore, some of the attributes of a successful music educator which a music major possesses may have to be pointed out to him before he applies for a position or exerts himself in such a way that his superiors may fire him. It seems, however, that intelligence would be the overriding factor helping the potential teacher to moderate his behaviors.

The personality and motivational profiles of the individual music education students could be used as guides in advising and counseling. If the student has a strong interest and desire in teaching music but is less experimenting and not group oriented, the institution would need to provide him with experiences to help him gain success in these areas. It is possible that a behavior modification program designed to affect changes in undergraduate music education students' behaviors could be successful. But, most importantly, the use of personality and motivational instruments such as the ones used in the study would identify individual strengths and weaknesses and help the adviser in structuring the best possible program for the individual student. In other words, advising and counseling would become more personalized.

The best type of program or model would be one which includes regression techniques for further validating the significances of the relationships. While point predictions could be made, the importance of the regression program would be to stabilize the equations, the relationships among personality, motivational, and criterion variables, at the teacher training institution by constantly cross validating and modifying as needed. While the student would not need to be rejected or selected, the regression approach would assist the adviser. For example, if one wanted to select potential music educators from the regression findings of the study, he could apply the b weights to individuals' scores on the personality and motivation tests and make a point prediction. The better technique and the one emphasized by Cronbach, Anastasi, and Tatuoka is to establish some type of minimal cutoff or selection ratio. The higher the selection cutoff, the lower the chances for making individual prediction errors from the data. Yet, in determining a cutoff score, attention should be given to the percentage of false rejects (those persons that were rejected that would have been successful) and to false positives (those persons that were accepted but were unsuccessful in the job) within the selected group.

> In certain situations, the cutoff point should be set sufficiently high to exclude all but a few possible failures. This would be the case when the job is of such a nature that a poorly qualified worker could cause serious loss or damage. An example would be a commercial airline pilot. Under other circumstances, it may be more important to admit as many qualified persons as possible, at the risk of

> including more failures. In the latter case, the number of false rejects can be reduced by the choice of a lower cutoff score. Other factors that normally determine the position of the cutoff score include the available personnel supply, the number of job openings, and the urgency or speed with which the openings must be filled (Anastasi 1968:133).

In advising and counseling, it would be unnecessary to set a cutoff point. Rather, the individual's score would be used in combination with other evidence to structure his program of studies and to assist him with the kinds of experiences best suited for helping him to become a successful music educator. The regression or discriminant approaches would satisfy the requirements for advising and counseling. Students would be neither accepted nor rejected on the bases of personality or motivation; but, they would be structured into learning experiences designed to develop their competencies for music teaching success. Personal and motivational competencies appear to be quite important and should be considered by teacher training institutions as part of the music teacher training program.

REFERENCES

Anastasi, Anne, Psychological Testing - Third Edition, New York: The Macmillan Company, 1968.

Campbell, Donald T. and Julian C. Stanley, "Experimental and Quasi-Experimental Designs for Research on Teaching," in Handbook of Research on Teaching, N. L. Gage, ed., Chicago: Rand McNally and Company, 1963, 171-246.

Cattell, Raymond B., Herbert W. Eber, and Maurice M. Tatsuoka, Handbook For The Sixteen Personality Factor Questionnaire (16PF), Champaign: Institute for Personality and Ability Testing, 1970.

Cronbach, Lee J. and G. C. Gleser, Psychological Tests and Personnel Decisions - Second Edition, Urbana: University of Illinois Press, 1965.

Tatsuoka, Maurice M., Discriminant Analysis: The Study of Group Differences, Champaign: The Institute for Personality and Ability Testing, 1970.

Tatsuoka, Maurice M., Validation Studies: The Use of Multiple Regression Equations, Champaign: The Institute for Personality and Ability Testing, 1969.

CHAPTER V

SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

The Problem

The purpose of the study was to investigate the relationships between personality and motivational variables and music teaching success criteria. Answers to the following questions provided partial answers to the problem:

1. What are the relationships among teacher accountability criteria used to define music teaching success:

   a. class mean residual gain scores of pupils
   b. teacher ratings by pupils
   c. teacher ratings by principals
   d. teacher ratings by supervisors
   e. teacher ratings by peers
   f. teacher self ratings

2. What are the relationships and interrelationships of the personality variables measured by the Sixteen Personality Factor Questionnaire and the motivational variables measured by the Motivation Analysis Test to music teaching success?

3. Do personality profiles differ for successful and unsuccessful male and female music educators?

4. What are the implications of the use of personality measures for counseling and advising undergraduate music education students?

The Sample

The study was divided into two main samples: music educator and music student teacher. The primary focus of the study was directed to the music educator sample which was partitioned as follows for analyses: (1) total music educator sample (N=209), (2) music educator derivation sample (80 percent of the total sample N=168), (3) music educator cross validation sample (20 percent of the total sample N=41), (4) female choral N=104, (5) female instrumental N=12, (6) male choral N=26, (7) male instrumental N=68, (8) female music educators N=115, and (9) male music educators N=94. These subjects were located in 16 states and 80 school districts in the United States and Puerto Rico.

A $15.00 gratuity was paid to 55 percent of the sample (N=115) in order to obtain subjects teaching music in socially disadvantaged schools and to attract teachers with widely varying abilities, years of teaching experience, and ages. The final sample included 63 percent of the subjects teaching in socially advantaged schools (N=131), 15 percent teaching in combination schools (N=31), and 22 percent (N=47) teaching in socially disadvantaged schools.

Teachers' data were not used in the final analyses if they failed to satisfy one or more of the following design requirements: (1) pre and post-testing of their pupils, (2) more than 12 pupils for music achievement testing, and (3) more than 2 rating evaluations. Subjects were also eliminated from the study if their data were suspect in any way.

From the total of 290 music educators originally selected from over 500 invited, 209 were used in the final analyses. The attrition from 290 to 209 subjects accounted for 28 percent of the total sample.

The music student teachers were 73 senior music education majors attending 7 teacher training institutions: University of Illinois, Eastern Illinois University, Western Illinois University, Illinois State University, Millikin University, Elmhurst College, and Oklahoma City University. From a population of over 183 music student teachers, 93 were selected to participate of whom 73 were used in the final analyses. Attrition from 93 to 73 subjects represented 22 percent of the music student teacher sample. Subjects were eliminated from the study if fewer than 3 rating evaluations were available.

The music student sample was partitioned as follows: female choral majors N=32, female instrumental majors N=12, male choral majors N=4, and male instrumental majors N=25. Of this sample, 64 percent of the subjects were attending large public teacher training institutions, 25 percent were attending medium sized public institutions, and 11 percent were attending medium sized private teacher training institutions. The primary use of the music student teacher sample was for cross validation purposes.

Music Teaching Success Criteria

Criterion validity was a principal concern in the study which sampled the domains of music teaching success. Six criteria were selected for investigation because these appeared to be major ways of evaluating teaching success:

1. Class mean residual gain scores of pupils
2. Teacher ratings by pupils
3. Teacher ratings by building principals
4. Teacher ratings by supervisors
5. Teacher ratings by peers
6. Teacher self ratings

These six criteria represented two primary domains of music teacher accountability: (1) an objective criterion represented by the class mean residual gain scores of pupils and (2) a subjective criterion represented by ratings of music teaching success.

Criterion referenced tests for the pupil gain construct were specially prepared for each participating music educator's class by the investigator from the standardized Music Achievement Tests I-IV by Richard J. Colwell. Permission was granted from the test publishers, the Follett Educational Corporation, to duplicate the machine scored answer sheets lithographically and to dub tape recordings from the published records. Music educators selected 5 music behavioral objectives from a list of 15 which related specifically to MAT subtests and pertained to what the music educators were teaching their pupils in the class used for music achievement testing. From the 5 selected objectives, the investigator chose 3 for pre and posttesting. A total of 47 specially prepared music achievement tests and tapes were prepared.

The 23,596 test answer sheets were scored by hand. Scoring accuracy was checked by the project consultant and found to be at 99 percent. The 1 percent error figure was substantially lower than the 28 percent average for hand scoring of standardized tests.

A multiple regression equation was used to derive the class mean residual gain score. Covariates included (1) a variable created from the following variables: (a) meetings times per week for each music class, (b) the minutes per session for each music class, (c) the interval for instruction between pre and posttesting, and (d) the socioeconomic classification of the school; (2) the meeting times per week variable; and (3) the class mean pretest score on the specially prepared music achievement test. By subtracting the actual class mean score from the predicted score, the deviation or class mean residual gain score was derived and became the G (gain) criterion. The sample was rank ordered from high to low in G.

The Illinois Teacher Evaluation Questionnaire (ITEQ) was used to gather the rating evaluations. ITEQ is the only published and standardized instrument designed for secondary school pupil evaluations of teachers although there are other published questionnaire designed for college students' evaluations of instruction and instructors. With permission from the publisher, the University of Illinois' Office of Measurement and Research, the investigator amended the 40 item ITEQ questionnaire to a 23 item form to facilitate the gathering of ratings from principals, supervisors, peers, and selves.

ITEQ is a forced choice questionnaire which uses a 4 point scale from strongly agree to strongly disagree for item statements. The music educator and music student teacher data were standardized for each sample from the 4 point scale to a ten point scale with a mean of 5.50 and a standard deviation of 2.00.

Missing ITEQ data were predicted for 66 of the 209 music educators and for 38 of the 73 music student teachers using multiple regression techniques. While 183 music educator subjects had complete ITEQ data and these data were used to generate the missing rating predictions, only 143 of these subjects were used in the final analyses because they satisfied the music achievement testing design requirements. Only 35 subjects in the music student teacher sample had complete rating data.

The 5 rating variables for both samples were then submitted separately to a "Principal Components-Principal Axis Factor Analysis Program" in which the investigator called for 5 factors from which the first principal component was used to generate a factor score for each participating subject. The weighted factor score was used as the R (rating) criterion for each sample. The samples were rank ordered from high to low in R.

Two additional criterion variables were constructed from the G and R domains for the music educator sample. G+R represented a composite music educator success criterion which added the R to the G score. G-R represented a discrepancy criterion which subtracted the R score from the G score. These criterion variables were 4 constructs of music teaching success: G (gain), R (rating), G+R (gain plus rating), and G-R (gain minus rating).

The Experimental Variables

The Sixteen Personality Factor Questionnaire (16PF) - 1967 Edition, Forms A and B by Raymond B. Cattell and Herbert W. Eber provided 16 experimental variables in the temperament trait domain of psychometric theory. These 16 factors retained their conceptual validities as unitary source traits in other cultures thus showing that they are of substantial scientific universality (Cattell and others 1970:13-14). Following are the 16 primary factors used in the study. The descriptive terms most meaningful to music educator interpretations are used rather than the clinical definitions.

BEST COPY AVAILABLE

A	Reserved, Detached	Outgoing, Participating
B	Low intelligence	High intelligence
C	Easily upset	Mature, Calm
E	Accommodating, Docile	Aggressive, Competitive, Stubborn
F	Sober, Serious, Taciturn	Enthusiastic, Happy-Go-Lucky
G	Expedient, Disregards rules	Conscientious, Persistent
H	Shy, Timid	Venturesome, Socially bold
I	Tough-minded, Realistic	Tender-minded, Sensitive
L	Trusting, Accepting	Suspicious, Hard to fool
M	Practical	Imaginative
N	Forthright, Unpretentious	Astute, Socially polished
O	Self Assured, Complacent	Apprehensive, Worrying, Insecure
Q1	Conservative, Traditional	Experimenting, Liberal
Q2	Group oriented, Joiner	Self sufficient, Resourceful
Q3	Lax, Careless of social rules	Socially precise, Controlled
Q4	Relaxed, Unfrustrated	Tense, Driven, Frustrated

The Motivation Analysis Test by Cattell and others provided 20 experimental motivation and interest variables in the dynamic dimensions of psychometric theory. MAT measured drives, feelings, sentiments, and attitudes in two areas, each including 10 variables: (1) Unintegrated Motive Strengths reflect that portion of the total dynamic energy that is unconscious and somewhat unfulfilled and (2) Integrated Motive Strengths reflect the conscious and satisfying dimension indicating that the individual has had successful experiences in those areas tapped by the test: Integrated Motive Strengths are like habit strengths. A conflict in motivation exists when the individual has a high Unintegrated (U) and a low Integrated (I) score on the same dimension. The 10 MAT U and I variables are listed below:

CA	Career Sentiment
HO	Home-Parental Sentiment
FR	Fear Drive
NA	Narcism-Comfort Drive
SE	Superego Sentiment
SS	Self Sentiment
MA	Mating Drive
PG	Pugnacity-Sadism Drive
AS	Assertiveness Drive
SW	Sweetheart-Spouse Sentiment

The above personality and motivational dimensions yielded 36 experimental variables for investigation: 16 from the 16PF test and 20 from the MAT test.

Because two main domains of psychometric theory were measured -- the personality trait domain and the dynamic domain -- analyses were conducted separately for each one: the 36 variables were not used together.

Although the tests were scored and normed so that both male and female subjects could be used in the same analyses (and most research studies in personality in music education combined the sexes), there was enough evidence in the clinical and counseling literature in personality to warrant investigating possible sex, personality-motivation, and criterion relationships. These analyses revealed the need for additional experimental moderating variables accounting for interaction effects. Moderators are those factors intervening between the predictor and criterion variable sets.

Sex and music teacher type or speciality (choral and instrumental) were included as two additional experimental variables. Furthermore, the analyses revealed differences between the sexes as the personality and motivational variables related to the music teaching success criteria.

For the 16PF temperament trait variables, interactions or differences between the sexes for music teaching success (e.g successful music educators) existed on Factors E, G, M, and N. This indicated that the male music educator tended to be more accommodating and the female more stubborn (E), the male more conscientious and the female more expedient (G), the male more imaginative and the female more practical (M), and the male astute and socially polished and the female more forthright.

One teacher type interaction was noted necessitating an additional experimental moderating type variable. Factor O when compared to teacher type and music teaching success criteria suggested that the successful instrumental music educator was more apprehensive, worrying, and insecure than the successful choral music educator who was more self assured and complacent.

Four sex and criterion interactions were found for the motivational variables: Unintegrated Home-Parental (HO-U), Unintegrated Self Sentiment (SS-U), Integrated Self Sentiment (SS-I), and Integrated Pugnacity-Sadism (PG-I). The successful male music educator had more unconscious interest in his home and family life than the successful female; the male was less social than the successful female who tended towards needing social and peer approval both consciously and unconsciously: she enjoyed socializing with others; and the male was more hostile and defensive than the successful female music educator. No teacher type and criterion interactions were discovered for the motivational variables.

A total of 11 additional experimental variables were included in the design of the study. Sex and teacher type were included as two dichotomous variables. In order to account for the interactions, the variables were multiplied times the interactionary cause in order to create moderators accounting for the interactions. The 9 interaction variables, moderators, were created from Factor E, Factor G, Factor M, Factor N, Factor O, HO-U, SS-U, SS-I, and PG-I.

> The most important test is between treatments, the first of the main effects. Next in importance, perhaps, equally important, are the interactions involving treatments. Take the interaction treatments X sex. If this were significant, it would mean that the amount of information a teacher possesses about students has an influence on student achievement, but boys are influenced differently than girls. Boys with teachers who possess information about their pupils may do better than boys whose teachers do not have such information, whereas it may be the opposite with girls, or it may make no difference one way or the other (Kerlinger 1973:356).

The above quote highlights the importance of investigating interactions and their effects. Because interactions needed to be handled, the sex and type variables at which interactions occurred were introduced as additional experimental variables created by multiplying the variable affected times sex or type thus creating the 9 additional experimental variables. Therefore, the personality trait analyses included 23 experimental variables (16 primary factors, sex, type, 4 moderating variables accounting for sex interactions, and 1 moderating variable accounting for type interaction) while the motivation analyses included 26 experimental variables (20 primary dimensions, sex, type, and 4 sex interacting moderating variables). While the 16PF test affords secondary and tertiary personality-trait dimensions, these are comprised of various combinations of primary factors -- the original 16 factors: these dimensions were not included as experimental variables because they would obfuscate the regression analyses.

The gathering of the personality and motivational data required between 3 and 4 hours of music educator and music student teacher questionnaire testing time. The 3 psychometric testing instruments yielded the most comprehensive personality trait and motivational measurement in music education research to date.

SUMMARY OF THE FINDINGS

Attrition

The reason frequently offered for refusing participation in the study was related to the considerable amount of music educator and pupil time perceived by potential subjects and their administrative superiors. This reduced the music educator sample from over 500 to 290 committed subjects and from over 183 to 93 music student teacher subjects. Affecting this attrition of the first type was perhaps an internalized fear of being objectively and subjectively evaluated in terms of pupil growth and ratings of music teaching success.

While both samples were weighted in favor of the female choral and male instrumental groups, these weightings were not markedly affected by secondary attrition: from 290 to 209 music educators (28 percent attrition) and from 93 to 73 music student teachers (22 percent attrition). Furthermore, attrition did not affect the samples as far as music educators teaching in large and small school districts or in socially advantaged to socially disadvantaged schools; and student teachers attending large public, small public, and small private teacher training institutions.

The music educators who were paid the $15.00 gratuity to participate in the study had a higher attrition rate than the unpaid subjects. This was perhaps due to a less internalized commitment to the study by the paid subjects.

The addition of the paid subjects partially helped to lower the bias of the unpaid subjects as far as commitment to the investigator, age, years of teaching experience, and amount of post bachelor's degree training were concerned. The paid subjects were older, had more years of teaching experience, and slightly less graduate training than the unpaid subjects.

A considerable amount of music student teacher data was collected directly from cooperating teachers and college supervisors. If these persons failed to complete the ratings or cooperate with the study, the student teachers data were not used. The music student teachers' fears of evaluation were not a factor in secondary attrition.

Attrition did not affect the relative weightings or biases of the samples: these remained relatively stable. It appeared that both samples were multifarious in nature.

Subproblem 1.

The first subproblem asked the question: what are the relationships among teacher accountability criteria used to define music teaching success:

a. class mean residual gain scores of pupils
b. teacher ratings by pupils
c. teacher ratings by principals
d. teacher ratings by supervisors
e. teacher ratings by peers
f. teacher self ratings

All the null hypotheses were tested beyond the .05 level of significance.

HO_1 The addition of the two covariates to the pretest class mean score will not significantly contribute to the residual gain score equation.

The created variable and the meeting times per week variable did not significantly contribute to the multiple regression equation. HO_1 could not be rejected, and the finding was that the residual gain program was similar to analysis of covariance because the class mean pretest score contributed the most to the prediction of the class mean posttest score.

> HO_2 There will be no significant differences among the 4 pretesting time groups in class mean residual gain scores (G).

A one by four analysis of variance found significant differences beyond the .01 level among the 4 pretesting time groups in G thus rejecting HO_2. t-test analyses found significant differences between groups 1 and 3 (September-October and January-February), groups 1 and 4 (September-October and March), and groups 1-2 and 3-4 (September-October, November-December and January-February and March) beyond the .01 level, and between groups 3 and 4 (January-February and March) beyond the .05 level. The music educator sample was stratified into 3 pretesting time groups: Group 1-2 reflecting pretesting times during the first semester months of September, October, November, and December; Group 3 reflecting pretesting during the first two months of the second semester (January and February); and Group 4 reflecting pretesting during March. The music educators were rank ordered from high to low according to class mean residual gain scores within each of the 3 pretesting groups. The total rank ordered sample was accomplished by percentage allocation according to the number of subjects in each pretesting time group. The subjects in each pretesting time group were assigned a rank according to the percentage of subjects in each group and were then rank ordered from 1 to 209.

> HO_3 There will be no significant intercorrletions among the 5 rating variables for the music educator sample.

HO_3 was rejected with the exception of the supervisor to self correlation. All of the other intercorrelations for the music educator ratings were significant beyond the .05 level; pupil to principal, pupil to supervisor, principal to peer, and principal to self; and supervisor to peer beyond the .01 level; and self to peer beyond the .05 level.

HO_4 There will be no significant intercorrelations among the 5 rating variables for the music student teacher sample.

HO_4 was rejected for the following music student teacher rating correlations: supervisor to cooperating teacher, supervisor to peer, and self to peer beyond the .01 level, and cooperating teacher to pupil ratings beyond the .05 level. HO_4 failed to be rejected for the following correlations: pupil to supervisor, pupil to peer, pupil to self, cooperating teacher to peer, cooperating teacher to self, and supervisor to self ratings.

HO_5 There will be no significant differences in rating evaluations among teacher type and grade level groups.

Analysis of variance analyses conducted on the 5 ratings of music teaching success for 6 groups of music educators (1. elementary choral, 2. elementary instrumental, 3. middle school choral, 4. middle school instrumental, 5. high school choral, and 6. high school instrumental) indicated no significant differences beyond the .05 level. HO_5 could not be rejected.

HO_6 There will be no identifiable rating factors for the music educator sample.

The factor analysis program which generated the factor scores for the R criterion also investigated the possibility of identifiable rating factors within the 5 rating variables. This was a byproduct of the study, and rating factors per se were not investigated in depth because no varimax rotation was used. The primary purpose was for weighting the ratings. Nevertheless, there were rating factors which appeared to emerge which indicated that HO_6 could be rejected even though the hypothesis was not tested at a predetermined level of significance. The first factor or possible subdomain appeared to be an adult rating factor as indicated by the closeness of the weights for the supervisor, principal, and peer rating variables: these 3 ratings appeared to cluster.

The second factor appeared to be the self rating domain while the third factor appeared to be the pupil rating subdomain. Factors 4 and 5 were less distinct but appeared to be peer and supervisor dimensions respectively.

> HO_7 There will be no identifiable rating factors for the music student teacher sample.

The principal components factor analysis which generated the factor rating scores for the music student teacher sample appeared to identify 4 possible rating subdomains. Like the music educator procedure, no varimax rotation was used although this holds interest for future investigations. Because there appeared to be 4 rating subdomains, HO_7 was tentatively rejected. The pupil dimension or subdomain appeared in Factor 2 while Factor 1 appeared to be an adult rating subdomain as indicated by the cooperating teacher and supervisor ratings. Factor 3 appeared to be the self rating dimension while Factors 4 and 5 appeared to be the cooperating teacher and supervisor dimensions respectively. Because of the possible invalidity of the music student teacher ratings, the above findings may be impuissant.

> HO_8 There will be no significant relationships between G and R for the following teacher type groups: female choral, female instrumental, male choral, and male instrumental.

Two analyses were conducted for testing HO_8. The first used the Spearman's rho technique for ranked data for investigating the relationships between G and R for the female choral, female instrumental, male choral, and male instrumental groups of the music educator sample. There were no significant relationships between G and R for the 4 teacher type groups as indicated by rho beyond the .05 level, and the HO_8 hypotheses failed to be rejected. When the regular Pearson product moment correlation coefficient was used to investigate HO_8, a slightly different finding occurred. The male instrumental group's R to G correlation was significant beyond the .10 level which, for the purposes of the study which accepted the .05 level, was not significant. The male choral group's R to G correlation was significant beyond the .05 level and the total male group's

R to G correlation was significant beyond the .02 level thus rejecting, in part, the null hypothesis. The correlational analyses for the female groups supported the rho findings. These correlations were barely positive in direction indicating that relationships existed but were minimal and insignificant. HO_8 was not rejected for the female groups and partially rejected for the male groups although these latter findings were somewhat disparate.

> HO_9 There will be no significant relationships between G and R for the music educator sample.

For the total sample, no significant relationship between G and R was found using the Pearson Product Moment Correlation. HO_9 was not rejected since there appeared to be no significant relationship between G and R for the total music educator sample.

> HO_{10} There were no significant relationships between G-R and R for total sample and teacher type groups.

The G-R to R correlations were all negatively significant beyond the .05 level. HO_{10} was rejected.

> HO_{11} There were no significant relationships between G-R and G for total sample and teacher type groups.

All the G-R to G correlations were significant beyond the .05 level. HO_{11} was rejected.

> HO_{12} There were no significant relationships between G+R and R for total sample and teacher type groups.

> HO_{13} There were no significant relationships between G+R and G for total sample and teacher type groups.

> HO_{14} There were no significant relationships between G+R and G-R for the total sample and teacher type groups.

All the G+R to R (HO_{12}) and G+R to G (HO_{13}) correlations were very close in actual numerical figures and were significant beyond the .01 level. The null hypotheses were rejected. The G+R to G-R correlations were all right around zero as expected for zero order correlations. HO_{14} failed to be rejected.

Hypotheses HO_{10} through HO_{14} substantiated the finding that two domains of music teaching success were being tapped in the study: an objective domain G and a subjective domain R. This finding required that separate regression analyses with the experimental variables would have to be conducted for each criterion.

HO_{15} There were no significant relationships among gain, the 5 rating variables, and sex for the total sample.

There were no significant correlations among gain, the 5 rating variables, and sex for the music educator sample. HO_{15} failed to be rejected further substantiating the finding of two music teaching criterion domains.

Subproblem 2

The second subproblem asked the question: what are the relationships and interrelationships of the personality variables measured by the Sixteen Personality Factor Questionnaire and the motivational variables measured by the Motivation Analysis Test to music teaching success?

HO_{16} There will be no significant interrelationships among 9 artificially stratified music teacher success groups, sex, and the 36 experimental personality and motivational variables.

The first analyses artificially stratified the total music educator sample into 9 success groups: (1) high rating-high gain, (2) high rating-medium gain, (3) high rating-low gain, and so forth to (9) low rating-low gain. Multivariate analysis of variance (MANOVA) was used to test hypothesis HO_{16} and included the 9 success groups, sex, and the 36 personality and motivational variables. A significant sex and experimental variables main effect was found beyond the .01 level, and HO_{16} was rejected.

HO_{17} There will be no perceptible sex interactions or interrelationships between the individual 36 experimental and the 4 criterion variables.

HO_{18} There will be no perceptible music teacher type interactions or interrelationships between the individual 36 experimental variables and the 4 criterion variables.

The next analyses generated the correlations between the predictors (the personality and motivational variables) and the 4 criteria for each teacher type sex group: female choral, female instrumental, male choral, and male instrumental. These correlational analyses aided in locating sex (HO_{17}) and type (HO_{18}) interactions experimental variable by variable since these appeared as differences among the teacher type subsamples in the predictor-criterion relationships. The interactions were located across all 4 criteria. While there were sometimes more pronounced differences for an experimental variable for certain groups on one or more of the 4 criteria, the identification of the interactions was made from differences in the correlations to the 4 criteria.

There appear to be two approaches for handling the problem of intervening variables or interactions. The first approach to this design problem is to create additional predictor variables which attempt to account for the interactions of sex and type. This allows for total sample predictions. This was done in Phase One of the study. The second is to stratify the sample into more homogeneous subgroups according to the interactions. This was done in Phase Two of the study which stratified the sample into music teacher sex type groups: female choral, female instrumental, male choral, and male instrumental.

Today, there is considerable controversy in the literature about the handling of intervening variables and the appropriate statistical models for use in research designs. Five statistical procedures include total group least squares linear regression, within group linear regression, Bayesian regression, full probability weighting, and reduced probability rating. All have their proponents and opponents, and no satisfactory solution has been rendered. The least squares linear regression models were used in the study.

> The more specific prediction models have the additional advantage of utilizing more information in predicting performance . . . [By] distinguishing subpopulations within the total sample, one is trading an increase in homogeneity for a loss in precision of parameter estimation. With respect to the present problem, the subgroup homogeneity has been well documented . . .

> A possible solution is the idea of incorporating the dimensions important in defining the subgroups as additional independent variables in the total least squares regression . . . The total least squares procedure . . . produced the optimal result . . . Clearly a direction for future research should involve examination of conditions which serve to recommend incorporation of intervening variables as additional predictors rather than using them to define subgroups (Lissitz and Schoenfeldt 1974:72).

Interactions necessitated the inclusion of additional experimental moderators in the prediction regression equations in order to account for the interaction. Otherwise, the ensuing predictions would result in obfuscations due to the interactionary effects. The personality and motivational variables which were suspected to be affected by interactions were multiplied times the interactionary cause, e.g. sex or teacher type, in order to create additional moderating experimental variables accounting for the interactions.

There appeared to be 8 sex interactions: Factor E, Factor G, Factor M, and Factor N from the 16PF variables and HO-U, SS-U, SS-I, and PG-I from the MAT variables. HO_{17} was rejected: there were perceptible sex interactions.

Only one type interaction appeared to be of sufficient magnitude to warrant the creation of one additional experimental moderating variable. This was 16PF Factor O. HO_{18} was rejected.

> There will be no significant relationships between the personality variables and the music teaching success criteria for the total music educator sample.

HO_{19} G criterion

HO_{20} R criterion

HO_{21} G+R criterion

HO_{22} G-R criterion

HO_{23} The above relationships, if significant, will not cross validate to a like sample of music educators.

The 16 primary factors from the 16PF as well as the 4 moderating sex and 1 moderating type variables as well as sex and type comprised the experimental variables for predicting music teaching success from the personality trait variables alone: HO_{19} through HO_{22} as well as for the cross validation hypotheses - HO_{23}. A total of 23 experimental variables was used to investigate the relationship between the personality variables and music teaching success criteria.

The multiple regression equation for the G criterion produced a multiple correlation coefficient of R=.26 which was significant beyond the .05 level. HO_{19} was initially rejected. However, on cross validation, the equation was not substantiated, and HO_{23} failed to be rejected for HO_{19}. The cross validation coefficient was r= -.08. The finding was that there appeared to be no significant relationships between the personality variables and the G criterion that were stable.

The equation for the R criterion (HO_{20}) produced a multiple R of .42 which was significant beyond the .05 level. HO_{20} was rejected. On cross validation, the equation was not significant (r= -.23) and HO_{23} for HO_{20} was not rejected. The finding was that while personality variables appeared to be related to R, these relationships were unstable.

The equation for the G+R criterion produced a multiple correlation coefficient of .39 which was significant beyond the .05 level. HO_{21} was initially rejected. The cross validation coefficient (r= -.13) was not significant, and HO_{23} for HO_{21} was not rejected. The finding was that while personality variables appeared to be related to G+R, the relationships were unstable.

No significant multiple regression equation was significant for the G-R criterion. HO_{22} failed to be rejected. The finding was that there appeared to be no relationships between the personality variables and the G-R criterion.

The overall finding from the above analyses was that the basic temperament trait domain as measured by the 16PF test was related to music teaching success criteria in terms of G, R, and G+R. On cross validation, however, the equations were not significant indicating cross validation sample instability.

This finding differs significantly from previous research.

> There will be no significant relationships between the motivational variables and the music teaching success criteria for the total music educator sample.

HO_{24} G criterion

HO_{25} R criterion

HO_{26} G+R criterion

HO_{27} G-R criterion

HO_{28} The above relationships, if significant, will not cross validate to a like sample of music educators.

Four sex criterion moderating variables plus sex and teacher type were added to the 20 experimental motivational variables for the regression analyses investigating relationships among motivational variables and music teaching success criteria. The multiple regression equation for the G criterion produced a multiple correlation coefficient of .35 which was significant beyond the .05 level which indicated that 12 percent of the variance in the G criterion could be accounted for by the motivational variables. HO_{24} was rejected. On cross validation, the equation was significant beyond the .01 level ($r= .40$), and HO_{28} for HO_{24} was rejected. The finding was that significant relationships existed between the motivational variables and the G criterion. The stability of the equation improved on cross validation and between 12 and 16 percent of the variance in G was accounted for by the motivational variables.

No significant multiple regression equations could be formulated to the R and G+R criteria. HO_{25} and HO_{26} as well as the cross validation equations HO_{28} were not rejected. The finding was that there appeared to be no relationships among the motivational variables and the R and G+R criteria.

The multiple regression equation for the G-R discrepancy criterion was significant beyond the .05 level. Cross validation was also significant. HO_{27} and the cross validation equation HO_{28} for HO_{27} were rejected. The finding was that relationships existed between the motivational variables and the G-R criterion. This may be the most significant finding of the study because it indicated that motivational variables help to explain the discrepancy between subjective and objective evaluations of a music educator's performance.

Phase Two

As a result of the Phase One findings discussed above, it was determined that sex and teacher type were interacting with the personality and motivational variables in fairly powerful ways. The net effect of such interactions was to lower the significance of the relationships for the combined sample; hence, the relatively low but significant multiple R's. Phase Two analyses attempted to see if the relationships among success criteria and the motivational and personality variables could be improved by stratifying the sample into teacher type sex groups.

Separate stepwise multiple regression equations were formulated for each teacher type sex group for the 16PF and MAT experimental variables including the 11 additional moderating, sex, and type variables for investigating the relationships between personality and motivational variables and the 4 music teaching success criterion variables. The total sample (N=209) was used because the derivation sample was too small when partitioned into the teacher type subgroups. As it was, there were too few subjects in the male choral and female instrumental subgroups to meaningfully interpret the data.

The music student teacher sample was used to cross validate the equations to the R criterion. Furthermore, the equations derived for each teacher type group were cross validated to the other teacher type groups. Heuristically this approach appeared sensible for investigating sample heterogeneity.

> There will be no significant relationships between the personality variables and the music teaching success criteria for female choral music educators.

HO_{29} G criterion

HO_{30} R criterion

HO_{31} G+R criterion

HO_{32} G-R criterion

HO_{33} The above relationships, if significant, will not cross validate to the music student teacher sample on the R criterion.

HO_{34} The above relationships, if significant, will not cross validate to the other teacher type groups.

There were significant relationships between the personality variables and the 4 criteria for the female choral group of music educators. HO_{29} through HO_{32} were rejected. The findings were that relationships existed: however, these equations did not cross validate to the other teacher type groups and to the music student teacher group for the R criterion. HO_{33} and HO_{34}, the cross validation hypotheses, failed to be rejected. Female choral music educators appeared to be homogeneous and different in personality from the other teacher type groups.

> There will be no significant relationships between the motivational variables and the music teaching success criteria for female choral music educators.

HO_{35} G criterion

HO_{36} R criterion

HO_{37} G+R criterion

HO_{38} G-R criterion

HO_{39} The above relationships, if significant, will not cross validate to the music student teacher sample on the R criterion.

HO_{40} The above relationships, if significant, will not cross to the other teacher type groups.

There were significant relationships between the motivational variables and the music teaching success criteria for female choral music educators.

HO_{35} through HO_{38} were rejected. These equations did not cross validate to the other teacher type groups nor to the music student teacher sample on the R criterion. HO_{39} and HO_{40} could not be rejected.

> There will be no significant relationships between the personality variables and the music teaching success criteria for male instrumental music educators.

HO_{41} G criterion

HO_{42} R criterion

HO_{43} G+R criterion

HO_{44} G-R criterion

HO_{45} The above relationships, if significant, will not cross validate to the music student teachers sample on the R criterion.

HO_{46} The above relationships, if significant, will not cross validate to the other teacher type groups.

There was a significant relationship between the personality variables and the R criterion for the male instrumental group. No relationships appeared to exist among the personality variables and the G, G+R, and G-R criteria. Cross validation to the music student teacher group on the R criterion was negative and not significant. HO_{49} was not rejected as was HO_{46}: there were no significant cross validations to the other teacher type groups. HO_{41}, HO_{43}, and HO_{44} failed to be rejected. The finding was that relationships appeared to exist between the personality variables and the R criterion: HO_{42} was rejected.

> There will be no significant relationships between the motivational variables and the music teaching success criteria for male instrumental music educators.

HO_{47} G criterion

HO_{48} R criterion

HO_{49} G+R criterion

HO_{50} G-R criterion

HO_{51} The above relationships, if significant, will not cross validate to the music student teacher sample on the R criterion.

HO_{52} The above relationships, if significant, will not cross validate to the other teacher type groups.

There were significant relationships between the motivational variables and the R, G+R, and G-R criteria. There was no significant relationship between the motivational variables and the G criterion for the male instrumental group. Cross validation was not significant. HO_{51} and HO_{52} failed to be rejected. HO_{48}, HO_{49}, and HO_{50} were rejected and indicated that relationships existed between the motivational variables and the R, G+R, and G-R criteria. HO_{47} failed to be rejected since no significant motivational equation predicted the G criterion. Because cross validation was insignificant, it appeared that the male instrumental group was homogeneous as far as personality and motivation were concerned.

In general, the significance of the relationships among personality, motivational and criterion variables was improved by partitioning the sample into music teacher type subgroups although these relationships could not be substantiated on like samples of subjects as they were in Phase One findings. The attempt to cross validate the equations to the other teacher type groups (internal cross validation) was to further investigate whether personality and motivational variables which were related to success for one subgroup were related to other teacher type subgroups for the same criteria. The finding that none of these cross validation null hypotheses could be rejected indicated that the music educator sample was heterogeneous. In other words, each teacher type subgroup was homogeneous requiring different combinations of personality and motivational variables which contributed to the multiple regression equations.

None of the significant equations for the personality and motivational variables to R cross validated to the music student teacher samples. Interesting, however, was that a hint of cross validation was evident for the female choral group in the personality domain. While the coefficient was not significant, there was possible evidence that with larger numbers of subjects and more valid criteria, the relationships might be enhanced.

There were significant relationships among personality and motivational variables and music teaching success criteria for the female choral and male instrumental groups. Further research will be needed to substantiate the relationships for the female instrumental and male choral groups as well as for cross validations.

Subproblem 3

The third subproblem asked the question: do personality profiles differ for successful and unsuccessful male and female music educators? Profiles were attempts to explain the differences between success groups of music educators. In essence, the investigator was searching for hints of symbiosis between the regression and the profile findings.

The music educator sample was stratified into male and female success groups for both G and R criteria: successful or high (upper 25 percent), average or middle (middle 50 percent), and unsuccessful or low (lower 25 percent). The average groups were not used in the analyses. Two group stepwise multiple discriminant analyses were used to test the profile hypotheses.

> There will be no significant personality profile differences between successful and unsuccessful groups of music educators for the G and R criteria when investigated separately by sex.

HO_{53} Male gain

HO_{54} Female gain

HO_{55} Male rating

HO_{56} Female rating

Null hypotheses HO_{53} through HO_{56} were rejected because there were significant personality profile differences between successful and unsuccessful groups of music educators beyond the .05 level of significance on the 4 criteria.

> There will be no significant motivational profile differences between successful and unsuccessful groups of music educators for the G and R criteria when investigated separately by sex.

HO_{57} Male gain

HO_{58} Female gain

HO_{59} Male rating

HO_{60} Female rating

Null hypotheses HO_{58} through HO_{60} were rejected beyond the .05 level of significance. HO_{57} failed to be rejected. There appeared to be significant motivational profile differences between successful and unsuccessful music educators for the female gain, male rating, and female rating criteria.

> There will be no significant personality profiles differences between male and female groups of music educators as stratified by success.

HO_{61} Successful gain

HO_{62} Unsuccessful gain

HO_{63} Successful rating

HO_{64} Unsuccessful rating

Null hypotheses HO_{61} through HO_{64} were rejected beyond the .05 level of significance. There were significant profile differences between successful and unsuccessful groups of male and female music educators.

> There will be no significant motivational profile differences between male and female groups of music educators as stratified by success.

HO_{65} Successful gain

HO_{66} Unsuccessful gain

HO_{67} Successful rating

HO_{68} Unsuccessful rating

Null hypotheses HO_{65} through HO_{68} were rejected beyond the .05 level. There were significant profile differences between successful and unsuccessful groups of male and female music educators.

BEST COPY AVAILABLE

The music educator who is successful in both gain and rating is intelligent, interested in social approval and reputation, is assertative and driven towards status, and is interested in the opposite sex, enjoys sensual indulgences of all kinds, and is somewhat hostile and defensive. The successful male music educator, in addition to the above, is stubborn, imaginative, apprehensive, socially precise, interested in and receives satisfaction from his home as well as his career, is interested in accomplishing things, social approval, and status. He appears to be extraverted.

The successful male in G is more shrewd and astute, experimenting, and interested in the opposite sex. The successfully rated male music educator tends to be more participating, enthusiastic, venturesome, concerned about status, and more ambitious. He appears traiditional and somewhat conservative.

The successful female music educator in G, in addition to the previously mentioned traits and motivations, is more stubborn and interested in accomplishing things. The successfully rated female, on the other hand, appears to be more suspicious, conscientious, shrewd and astute, experimenting, and socially precise. All music educators appear interested in sensual indulgences. Males are more imaginative, intelligent, and creative than the females.

The above discussion of the profile findings of the study highlighted some of the similarities and differences among the successful and unsuccessful groups of male and female music educators on the G and R criteria. The profile data assisted in the interpretation of the significance of the relationships of personality and motivational variables to music teaching success.

Subproblem 4

The fourth subproblem asked the question: what are the implications of the use of personality measures for counseling and advising undergraduate music education students? The answer to this question was not found by rigorously testing hypotheses but rather by the process of induction: the specific findings of the study were used to generate counseling and advising principles.

The major goal of the study was to investigate the relationships among personality, motivational, and criterion variables in order to counsel and advise prospective music teachers. The regression and profile analyses were conterminous techniques. The study of relationships is apposite to counseling and advising because before one can advise and counsel future music educators, he must first identify the personality and motivational dimensions related to music teaching success.

The findings of the study indicated that differences in personality and motivation exist for successful and unsuccessful male and female music educators as well as for music teacher type groups. The regression analyses and the profile data confirm some of the usefulness of the personality and motivational information as far as advising and counseling music education majors is concerned.

CONCLUSIONS

The findings of the regression and profile discriminant analyses led the investigator to conclude that personality and motivational factors are related to music teaching success in fairly powerful ways. The validity of these relationships is affected by (1) the criteria chosen by the researcher to define music teaching success, (2) the reliability and validity of the data gathering instruments, (3) the heterogeneity of the sample selected for investigation, and (4) the identification and control of the intervening variables between experimental and criterion variable sets.

It appears that researchers in selecting criteria for evaluating music teaching success will have to adopt a philosophical basis as to the kinds of data to be gathered. No single criterion of music teaching success appears valid enough to warrant merit pay, advancement, or tenure judgments. Philosophical decisions must be made by those having vested interests in the successes of music teachers in relation to pupil expectancy areas about the types of measures used to gather and weight accountability criteria. Process, product, and opinion categories are important. Investigating relationships among personality, motivational, and criterion variables and varying linear combinations or profiles will emerge depending upon the evaluative evidence. Future researchers will need to assume that the criterion domain is multivariate in nature and design studies accordingly.

To the investigator's knowledge, no study to date has found a significant relationship between objective measures of teaching success in terms of gain scores and subjective measures of success such as ratings. The findings of the present study confirm those of Hellfritzsch (1945), Linx (1946), Barr (1935, 1953), Morsh (1956), and Heil and others (1960). Furthermore, the variability in the reliability of ratings as found by Lutz (1963) between pupil and administrator ratings was affirmed in that the factor analysis of the ratings in the present study suggested that the raters were rating the music educators from different frames of reference. The findings as reported by Silberman (1971) that pupil ratings of student teachers were not significantly related to college supervisor ratings was somewhat affirmed due to the possibilities of rating subdomains. Ratings must be questioned as evaluation and accountability criteria.

Personality trait instruments, while not as reliable and valid as achievement tests, are useful in gathering data for differentiating among music educators. Motivations are expected to change as individuals mature and grow in their feelings, sentiments, and drives. The Motivation Analysis Test is a useful instrument not only for investigating relationships between motivational variables and music teaching success but for advising and counseling prospective music teachers. Since feelings and drives are more modifiable than basic personality traits, motivational experiences should be included in the music teacher training curriculum. Furthermore, personality and motivational dimensions appear to be interrelated to some extent necessitating that the adviser consider the total picture or profile and use multivariate techniques for providing this information.

The heterogeneity of the music educator sample affected the relationships among the personality and motivational variables and music teaching success criteria. Music educators are heterogeneous as far as personality and motivation are concerned even though previous investigations by Bodenstab (1968), Funk (1973), and with music student teachers, Wink (1967), found or assumed homogeneity. Even though music educators are psychologically homogeneous in contrast to other classroom teachers and administrators, sex and music teaching speciality appear to be of sufficient importance that partitioning of music educator and music student teacher samples into teacher type or major and sex groups appears warranted.

Any research investigation that investigates the relationships among experimental and criterion variables should include means for further substantiating these relationships and findings. One of the best techniques for substantiation is cross validation: applying the findings to another analogous sample of subjects. It is accepted today in the psychometric literature that multiple regression equations and discriminant analyses should be cross validated.

Correlations between the experimental and criterion variables which appeared originally to be important did not necessarily remain significant in the regression and discriminant analyses. This indicates that previous procedures for selecting variables for inclusion in equations are questionable. Simple correlations among experimental and criterion variables or t-test and chi square techniques between success groups on each experimental variable which produced significant findings in other studies are of little value for investigating relationships or differences among groups for advising and counseling purposes.

Individual variable procedures do not account for significant interactions and interrelationships which affect the nature of the findings.

The fact that between 12 and 17 percent of a music teacher's success in terms of G and G-R as measured by the instruments used in the study can be accounted for by motivational variables was a major finding. Furthermore, the findings suggest that with more subjects in the teacher type subgroups, personality trait variables will be more significantly related to music teaching success. The fact that the significance of the relationships was improved by stratifying the sample into male and female, instrumental and choral groups, as well as stratifying for the profile analyses, indicated that personality and motivational variables should be used in advising and counseling potential and inservice music teachers as long as these subjects are, also, partitioned similarly.

Interpersonal relationships seemed to be important to differentiating among music educator success groups. The rating criterion showed that the successful music educator is more traditional, serious and accommodating, less experimenting and assertive, and more attached to the home. This person would not "rock the boat" and irritate those in charge. The gain criterion, on the other hand, showed that the successful music educator is more participating, experimenting, enthusiastic, dominant and stubborn, satisfied in music teaching as a career, interested in the opposite sex, is defensive, and interested in achieving success with strong feelings toward the home, and is interested in his general reputation. He is anxious, creative, imaginative, and extraverted. The successful music educator is not afraid of doing his own thing and has high interests in sensual indulgences. The successful male is more socially aware and outgoing while the female is more self sufficient and resourceful.

Principles for counseling and advising undergraduate music education students were inferred from the findings of the study.

1. The sexes must be investigated separately. A multiple regression or discriminant approach to advising and counseling should investigate the relationships separately for the sexes. Intervening relationships should be treated by incorporating moderators in the equations. Preferably, music teacher type and sex subgroups would be studied: female choral, female instrumental, male choral, and male instrumental. The regression or discriminant models would be used to investigate relationships among experimental and criterion variables and differences among groups, not for the purpose of selection and rejection, but for assisting the student to select the best possible experiences within the music teacher training program.

2. A student who is lacking in certain areas found to be important to music teaching success would be provided the opportunity to gain confidence and security by a sequential program of experiences which would assist his development of competencies. A student who is introverted may not be changed to an extravert over night, but he will be provided the opportunity to gradually develop confidence in his relationships with other persons: adults, peers, and pupils. A student who is interested primarily in working alone will be provided opportunities to experience group activities and group successes.

3. The student will learn how to relate to persons who do not care for the things he cherishes or act the way he acts. He will understand that his interests in self indulgences are normal but that others may not consider them to be so, and he will need to learn how to deal with these persons. Stubborness, while important to music teaching success, may not provide the future music educator with the ammunition needed to be successful the first few months on the job especially with his immediate superiors.

4. The music teacher training curriculum will assist the student to develop his personality and motivations and to modify those aspects which are not conducive to his potential success. While basic personality traits are more

stable and unchanging, motivations change. The curriculum as implemented by the teaching staff should be insouciant to the personal and motivational needs of the music student.

5. Advising and counseling would be personal. One to one communication between student and professor will provide the best climate for student success as the teacher training program becomes a means to developing music teaching success potentialities and not an end in itself.

Music student teachers' scores did not serve for cross validation purposes as originally desired. Some evidence supports the conclusion that criteria for music teaching success and music student teaching success are different. The same rating form was used, and the 5 rating categories were approximately similar. Yet, the findings related to the student teaching sample are in doubt due to criterion validity questions. It may be necessary that music student teaching criteria include some type of teaching task objective evaluative instruments along the lines presently under investigation by Popham and others.

RECOMMENDATIONS

Replication research is not only admirable, it is obligatory for studies investigating relationships among experimental and criterion variables. Furthermore, multiple regression equations and discriminant functions need to be continuously validated on new subjects in local situations. Even though two of the major findings of the study were significantly cross validated in no way implies that these data are appropriate to other groups of music educators. The following specific recommendations are offered:

1. The personality trait variables measured by the 16PF Test and the motivational variables measured by the MAT Test should be used in a multivariate design with larger numbers of subjects in the teacher type sex groups in order to facilitate cross validation procedures. Variables relating to the music teaching situation, the socioeconomic nature of the school, the size of the

school district, and the ages, years of teaching experience, and post bachelor's degree training of subjects should be investigated in order to determine how these variables relate to music teaching success criteria.

2. The research design should include all the experimental variables rather than treating the personality and motivational variables as two separate psychometric domains. There is some evidence to suggest that motivational variables may interact with the personality trait variables in such a way as to affect better interpretations of the data.

3. Intervening variables sets between the experimental and criterion variables should be investigated. There is evidence to suggest that these interactions affect the relationships. Sex and teacher type were found important in the present study, but this does not exclude the possibility of other factors such as age, music teaching situation, grade level, and so forth.

The criterion domain of music teaching success is elusive. While ratings have been the most common criteria used in previous research investigations, other factors appear important such as pupil attitudinal variables, pupil process and product variables, and music teacher attitudinal variables. Not everything in music teaching is measurable. If research investigations concentrate on success criteria that are explicit such as pupil gain measures, much information may be lost; and, furthermore, the teacher may be teaching only measurable behaviors while neglecting the important affective behaviors.

4. Investigations in the criterion domain of music teaching success are urgently needed. One possible approach would be to include as many variables as possible and investigate the data using canonical correlation techniques. With this approach, one might be able to draw some inferences about relationships and interrelationships.

5. The criteria for evaluating music student teacher success should be investigated in relation to music teaching success criteria. What is a successful music student teacher, and how does this affect his actual music teaching

success? Some longitudinal investigations in the criterion domain appear needed.

Perhaps Maslow best defined the type of person required for success in his chosen vocation. Research in music education might well direct itself to investigating the satisfied and competent music educator.

> In examining self-actualizing people directly, I find that in all cases, at least in our culture, they are dedicated people, devoted to some task "outside themselves," some vocation, or duty, or beloved job . . . One gets the feeling of a beloved job, and, furthermore, of something for which the person is a "natural," something that he is suited for, something that is right for him, even something that he was born for . . . In the best instances, the person and his job fit together and belong together perfectly like a key and lock, or perhaps resonate together like a sung note which sets into sympathetic resonance a particular string in the piano keyboard (Maslow 1971:301-302).

Finally, the end result of all research investigation must benefit children. Not only are their cognitive and psychomotor learnings important but the processes leading to affective and aesthetic sensitivities help to develop the self actualizing human being -- the end goal for which all strive throughout their lives. The development of outstanding music educators, sensitive to children and their needs, must include personality and motivational research.

REFERENCES

Cattell, Raymond B., Herbert W. Eber, and Maurice M. Tatsuoka, Handbook for the Sixteen Personality Factor Questionnaire (16PF), Champaign: Institute for Personality and Ability Testing, 1970.

Cattell, Raymond B., and J. J. Butcher, The Prediction of Achievement and Creativity, Indianapolis: Bobbs-Merrill, 1968.

Kerlinger, Fred N., Foundations of Behavioral Research - Second Edition, New York: Holt, Rinehart, and Winston, 1973.

Lissitz, Robert W., and Lyle F. Schoenfeldt, "Moderator Subgroups for the Estimation of Educational Performance: A Comparison of Prediction Models," American Educational Research Journal, Winter 1974: II, 1, 63-75.

Maslow, Abraham H., The Farther Reaches of Human Nature, New York: The Viking Press, 1971.

BEST COPY AVAILABLE

BIBLIOGRAPHY

BOOKS

Anastasi, Ann. Psychological Testing - Third Edition. Toronto: The Macmillan Company, 1968.

Ausebel, David P. Educational Psychology. New York: Holt, Rinehard and Winston, 1968.

Barr, Arvil S., Robert A. David, and Palmer O. Johnson. Educational Research and Appraisal. New York: J. B. Lippincott Company, 1953.

Barr, A. S., William H. Burton, and Lee J. Brueckner. Supervision: Principle and Practices in the Improvement of Instruction. New York: A. Appleton Century Company, 1938.

Bloom, Benjamin S., J. Thomas Hastings, and George F. Madaus. Handbook On Formative and Summative Evaluation of Student Learning. New York: McGraw-Hill Book Company, 1971.

Cattell, Raymond B., and John L. Horn. Handbook for the Motivation Analysis Test (MAT). Champaign: Institute for Personality and Ability Testing, 1964.

Cattell, Raymond B., Herbert W. Eber, and Maurice M. Tatsuoka. Handbook for the Sixteen Personality Factor Questionnaire (16PF). Champaign: Institute for Personality and Ability Testing, 1970.

Cattell, Raymond B., and J. J. Butcher. The Prediction of Achievement and Creativity. Indianapolis: Bobbs-Merrill, 1968.

Cattell, Raymond B. The Scientific Analysis of Personality. Baltimore: Penquin, 1965.

Colwell, Richard J. The Evaluation of Music Teaching and Learning. Englewood Cliffs; Prentice-Hall, Inc., 1970.

Colwell, Richard J. Interpretive Manual - Music Achievement Tests I and II. Chicago: Follett, 1969.

Cronbach, Lee J. and G. C. Gleser. Psychological Tests and Personnel Decision - Second Edition. Urbana: University of Illinois Press, 1965.

Dewey, John. Democracy and Education. New York: Free Press, 1916.

Flanders, Ned A. Analyzing Teaching Behavior. Reading: Addison-Wesley Publishing Company, 1970.

Glass, Gene V., and Julian C. Stanley. Statistical Methods in Education and Psychology. Englewood Cliffs: Prentice-Hall, Inc., 1970.

Graffam, Donald T. Dickinson College Changes Personality. Dickinson College, 1970.

Guilford, J. P. Psychometric Methods. New York: McGraw Hill, 1954.

Heil, Louis M., Marion Powell and Irwin Feifer. Characteristics of Teacher Behavior and Competency Related to the Achievement of Different Kinds of Children in Several Elementary Grades. Washington: U. S. Office of Education, Cooperative Research, SAE 7285.

Hoffer, Charles R. Teaching Music in the Secondary Schools - Second Edition. Belmont: Wadsworth Publishing Company, Inc., 1973.

Kerlinger, Fred N. Foundations of Behavioral Research - Second Edition. New York: Holt, Rinehart and Winston, Inc., 1973.

Klotman, Robert H. The School Music Administrator and Supervisor: Catalysts for Change in Music Education. Englewood Cliffs: Prentice-Hall, Inc., 1973.

Lehman, Paul R. The Seventh Mental Measurements Yearbook, Volume I. Oscar K. Buros, ed. Highland Park: The Gryphon Press, 1972.

Lehman, Paul R. Tests and Measurements in Music. Prentice-Hall, Inc. 1968.

Leonhard, Charles and Robert W. House. Foundations and Principles of Music Education - Second Edition. New York: McGraw Hill Book Company, 1972.

Lucio, William H. and John D. McNeil. Supervision: A Synthesis of Thought and Action - Second Edition. New York: McGraw-Hill Book Company, 1969.

McCall, William A. How To Measure In Education. New York: The Macmillan Company, 1922.

Madsen, Clifford K. and Charles H. Madsen, Jr. Experimental Research in Music. Englewood Cliffs: Prentice-Hall, Inc., 1970.

Mahan, J. Maurice. Manual of Interpretation for the Illinois Teacher Evaluation Questionnaire. Champaign: Measurement and Research Division, Office of Instructional Resources, June, 1970.

Maslow, Abraham H. The Farther Reaches of Human Nature. New York: The Viking Press, 1971.

Morgan, Russell Van Dyke and Hazel Nohavec Morgan. Music Education in Action. Park Ridge: Neil A. Kjos Music Company Publishers, 1960.

Moses, Harry E. Developing and Administering a Comprehensive High School Music Program. West Nyack: Parker Publishing Company, Inc., 1970.

Music Educators National Conference. Teacher Education In Music: Final Report. Washington: Music Educators National Conference, 4, 1972.

National Council of State Supervisors of Music. Guidelines in Music Education: Supportive Requirements. Washington: Music Educators National Conference, 9, 1972.

Reimer, Bennett. A Philosophy of Music Education. Englewood Cliffs: Prentice-Hall, Inc., 1970.

Ryans, David G. Characteristics of Teachers. Washington: American Council on Education, 1960.

Schneider, Erwin H. and Henry L. Cady. Evaluation and Synthesis of Research Studies Relating to Music Education. Washington: U. S. Office of Education, Cooperative Research, E-016, The Ohio State University, 1965.

Silberman, Charles E. Crisis in the Classroom: The Remaking of American Education. New York: Random House, 1970.

Tatsuoka, Maurice M. Discriminant Analysis: The Study of Group Differences. Champaign: Institute for Personality and Ability Testing, 1970.

Tatsuoka, Maurice M. Multivariate Analysis: Techniques for Educational and Psychological Research. New York: John Wiley and Sons, Inc., 1971.

Tatsuoka, Maurice M. Validation Studies: The Use of Multiple Regression Equations. Champaign: The Institute for Personality and Ability Testing, 1969.

Whybrew, William. Measurement and Evaluation in Music - Second Edition. Dubuque: Wm. C. Brown, Co., 1971.

PERIODICALS

Albert, H. R. "An Analysis of Teacher Rating by Pupils in San Antonio, Texas." Educational Administration and Supervision, XVI, January 1941.

Aleamoni, Lawrence M. "A Review of Recent Reliability and Validity Studies on the Illinois Course Evaluation Questionnaire (CEQ)." Research Memorandum #127, May 1972.

Aleamoni, Lawrence M. "Illinois Course Evaluation Questionnaire (CEQ) Results Interpretation Manual Form 66 and Form 32." Research Report #331, 1972.

Anderson, Harold M. "A Study of Certain Criteria of Teaching Effectiveness." Journal of Experimental Education, XXIII, September, 1954.

Barr, Arvil S. "Characteristics of Successful Teachers." Phi Delta Kappan, March 1958.

Barr, Arvil S. "The Measurement and Prediction of Teaching Efficiency." Review of Educational Research, June 1958.

Barr, Arvil S. "The Measurement and Prediction of Teaching Efficiency: A Summary of Investigations." Journal of Experimental Education, 1961.

Barr, Arvil S. "Measurement and Prediction of Teaching Success." Review of Educational Research, 1949.

Barr, Arvil S. "The Measurement of Teacher Characteristics and Prediction of Teaching Efficiency." Review of Educational Research, 1962.

Barr, Arvil S. "Problems Associated With The Measurement and Prediction of Teaching Success." Journal of Educational Research, 1958.

Barr, Arvil S., Theodore L. Torgerson, Carl E. Johnson, Vergil E. Lyon, and Anthony C. Walvoord. "The Validity of Certain Instruments Employed in the Measurement of Teaching Ability." The Measurement of Teaching Efficiency, ed. Helen M. Walker. New York: The Macmillan Company, 1935.

Bereiter, Carl E. "Some Persisting Dilemmas in the Measurement of Change." Problems in Measuring Change. Chester W. Harris, ed. Madison: University of Wisconsin Press, 1963.

Betts, Gilbert L. "Pupil Achievement and the NS Trait in Teachers." The Measurement of Teaching Efficiency. Helen M. Walker, ed. New York: The Macmillan Company, 1935.

Bodenstab, Ernest W. "Personality Patterns and Heterogeneity Among Teaching Interest Areas: A Multivariate Approach." AERA Paper Abstracts, 1968 New York Annual Meeting.

Borkowski, Francis Thomas. "The Relationship of Quality of Work in Undergraduate Music Curricula to Effectiveness of Instrumental Music Teaching in the Public Schools." Journal of Experimental Education, Fall, 1970.

Brophy, Jere E. "Stability of Teacher Effectiveness." American Educational Research Journal. Summer, 1973.

Bruner, Jerome S. "Process of Education Revisited." Phi Delta Kappan, September, 1971

Cady, Henry L. "A Conference on Research in Music Education." Council for Research in Music Education Bulletin. Fall, 1969.

Cady, Henry L. "Tests and Measures in Higher Education: School Music Teachers." Journal of Research in Music Education. XV, 1967.

Campbell, Donald T. and Julian C. Stanley. "Experimental and Quasi-Experimental Designs for Research on Teaching." Handbook of Research on Teaching. N. L. Gage, ed. Chicago: Rand McNally Company, 1963.

Cass, James. "Accountable To Whom? For What?" Saturday Review. March 20, 1971.

Cattell, Raymond B. "Personality Measurement Functionally Related To Source Trait Structure." Measurement in Personality and Cognition. Messick & Ross, eds. New York: Wiley, 1962.

Cattell, Raymond B., K. Barton, and T. E. Dielman. "Prediction of School Adhievement from Motivation, Personality, and Ability Measures." Psychological Reports, 1972.

Centra, John A. "Do Student Ratings of Teachers Improve Instruction?" Change Magazine. April, 1973.

Charters, W. W. Jr. "Survival in the Profession: A Criterion For Selecting Teacher Trainees." Journal of Teacher Education. September, 1956.

Colwell, Richard J. "Music Education and Experimental Research." Journal of Research in Music Education, XV, 1967.

Crews, E. Katherine. "The Role of the Supervisor of Student Teaching in Music - A Critique." Council for Research in Music Education Bulletin. Spring, 1970.

Cronbach, Lee J. and Lita Furby. "How We Should Measure 'Change'--Or Should We?" Psychological Bulletin, 1970.

Dayton, C. Mitchell. "The t-Test: Parametric of Non-Parametric." 1967 AERA Annual Report.

Durflinger, Glenn W. "Personality Correlates of Success in Student Teaching." Educational and Psychological Measurement, XXIII, 1963.

Dyer, J. S., R. L. Linn, and M. J. Patton. "A Comparison of Four Methods of Obtaining Discrepancy Measures Based on Observed and Predicted School Systems Means on Achievement Tests." American Educational Research Journal, 1969.

Ehlert, Jackson K. "Causes for Failures Among Music Teachers." Music Educators Journal. January, 1951.

Erickson, Harley E. "A Factorial Study of Teaching Ability," Journal of Experimental Education, XXIII, 1954.

Ernest, David J. "The Prediction of Academic Success of College Music Majors." Journal of Research in Music Education, XVIII, Fall, 1970.

Farwell, Gaylord H. "Factors Influencing Male Music Education Graduates of Certain Pennsylvania Institutions of Higher Education to Leave or Not Enter the Profession - A Critique." Council for Research in Music Education Bulletin, Spring, 1964.

Flanagan, Carroll Edward. "The Predictive Value of the Minnesota Multiphasic Inventory." Journal of Experimental Education. XXIX, June, 1961.

Fosse, John B. Critique - "The Relationship of Pre-Service Measures to Ratings of Music Teaching Success." Charles H. Benner, Council for Research in Music Education Bulletin. Spring, 1966.

Getzels, J. W. and P. W. Jackson. "The Teacher's Personality and Characterisitcs." Handbook of Research on Teaching. N. L. Gage, ed. Chicago: Rand McNally Company, 1963.

Gordon, Edwin. "Third Year Results of a Five Year Longitudinal Study of the Musical Achievement of Culturally Disadvantaged Students." Experimental Research in the Psychology of Music - Volume 8. Edwin Gordon, ed. Iowa City: University of Iowa Press, 1972.

Hellfritzsch, A. G. "A Factor Analysis of Teaching Abilities." Journal of Experimental Education, XIV, December, 1945.

Hill, John D. "A Study of the Musical Achievement of Culturally Deprived Children and Culturally Advantaged Children at the Elementary School Level." Experimental Research in the Psychology of Music - Volume 6. Edwin Gordon, ed. Iowa City: University of Iowa Press, 1970.

Holland, I. I. and R. I. Beazley. "Personality, Motivation and Education Needed in Professional Forestry." Journal of Forestry, July, 1971.

Holland, James G. and Judith Doran. "Instrumentation of Research in Teaching." Second Handbook of Research on Teaching. Robert M. W. Travers, ed. Chicago: Rand McNally Company, 1973.

Jones, Ronald D. "The Prediction of Teaching Efficiency from Objective Measures." Journal of Experimental Education, XV, 1946.

Killackey, Jim. "Accountability Roots Planted." The Sunday Oklahoman, September 9, 1973.

Lamke, Tom A. "Personality and Teaching Success." Journal of Experimental Education, XX, December, 1951.

Lee, Robert E. "An Investigation of the Use of the Musical Aptitude Profile with College and University Freshman Music Students." Journal of Research in Music Ecucation, XV, 4, 1967.

Lehman, Paul R. "The Predictive Measurement of Musical Success." Journal of Research in Music Education, XVII, 1, 1961.

Lehman, Paul R. "A Selected Bibliography of Works on Music Testing." Journal of Research in Music Education, XVII, 4, 1969.

Lehman, Paul R. and others. "School Personnel Research and Evaluation Services; Common Examinations; Music Education Examinations." a critique in the Journal of Research in Music Education, 20, 3, 1972.

Light, Richard J. "Issues in the Analysis of Qualitative Data." Second Handbook of Research on Teaching. Robert M. W. Travers, ed. Chicago: Rand McNally Company.

Lins, Leo Joseph. "The Prediction of Teaching Efficiency." Journal of Experimental Education, XV, 1946.

Lissitz, Robert W. and Lyle F. Schoenfeldt. "Moderator Subgroups for the Estimation of Educational Performance: A Comparison of Prediction Models." American Educational Research Journal, II, 1, Winter, 1974.

Lord, Frederic M. "Elementary Models for Measuring Change." Problems in Measuring Change. Chester W. Harris, ed. Madison: University of Wisconsin Press, 1963.

Lord, Frederic M. "A Paradox in the Interpretation of Group Comparison." Psychological Bulletin, 68, 1967.

Lord, Frederic M. "Statistical Adjustments When Comparing Pre-existing Groups." Psychological Bulletin, 72, 1969.

McNeil, John D. and W. James Popham. "The Assessment of Teacher Competence." Second Handbook of Research on Teaching. Robert M. W. Travers, ed. Chicago: Rand McNally Company, 1973.

Marks, Edmond and Charles G. Martin. "Further Comments Relating to the Measurement of Change." American Educational Research Journal 10, 3, 1973.

Michael, William B., George Barth, and Henry F. Kaiser. "Dimensions of Temperament in Three Groups of Music Teachers." Psychological Reports, 9, 1961.

Mitzel, Harold E. "Teacher Effectiveness." Encyclopedia of Educational Research - Third Edition. Chester W. Harris, ed. New York: The Macmillan Company, 1960.

Morsh, Joseph E., et al. "Student Achievement as a Measure of Instructor Effectiveness." Journal of Educational Psychology, XLVII, February, 1956.

Music Educators Journal, May, 1973. "NEA Reports."

Music Educators National Conference, 48.

Nunnally, Jim. "The Analysis of Profile Data." Psychologcal Bulletin, 59, 4, 1962.

O'Connor, Edward F. Jr. "Extending Classical Test Theory To The Measurement of Change." Review of Educational Research, 42, 1, 1972.

Popham, W. J. "Performance Tests of Teaching Proficiency: Rationale, Development, Validation." American Educational Research Journal, 8, 1971.

Prince, Warren. "Fulfilling the Need for Replication in Music Education Research." Council for Research in Music Education Bulletin, 20, Spring, 1970.

Rainbow, Edward. "Patterns of Teacher-Student Interaction in Selected Junior High School General Music Classes - A Critique." The Council for Research in Music Education Bulletin, 26, Fall, 1971.

Rawls, Donna J. and James R. Rawls. "Personality Characteristics and Personal History Data of Successful and Less Successful Executives." Psychological Reports, 23, 1968.

Schwartz, Anthony N. "A Study of the Discriminatory Efficiency of Certain Tests of the Primary Source Personality Traits of Teachers." Journal of Experimental Education, XIX, September, 1950.

Shatin, Leo. "Personality Traits of Music Therapists." Psychological Reports, 1968.

Sidnell, Robert G. "A Comparative Study of the Musical Achievement of Students in Three Illinois High Schools - A Critique." Council for Research in Music Education Bulletin, 16, 1971.

Sidnell, Robert G. "The Dimensions of Research in Music Education." Council for Research in Music Education Bulletin, 29, 1972.

Simpson, Edwin L. Critique = "The Relationship of Quality of Work in Undergraduate Music Curricula to Effectiveness of Instrumental Music Teaching in the Public Schools." Francis Borkowski, Council for Research in Music Education Bulletin, 16.

Singer, Arthur. "Social Competence and Success in Teaching." Journal of Experimental Education, XXIII, December, 1954.

Tatsuoka, Maurice M. and David V. Tiedeman. "Statistics as an Aspect of Scientific Method in Research on Teaching." Handbook of Research on Teaching. N. L. Gage, ed. Chicago: Rand McNally Company, 1963.

Taylor, Corwin H. "Opinions of Music Teachers Regarding Professional Preparatory in Music Education." Journal of Research in Music Education, XVIII, v, 1970.

Whellams, Frederick S. "The Relative Efficiency of Aural-Musical and Non-Musical Tests as Predictors of Achievement in Instrumental Music," Council for Research in Music Education Bulletin, 21, Summer, 1970.

White, Howard G. "Incidence and Causes of Withdrawal from Public School Music Teaching - A Critique." Council for Research in Music Education Bulletin, 6, Fall, 1965.

White, Howard G. "The Professional Role and Status of Music Educators in the United States." Journal of Research in Music Education, XV, 1, 1967.

Wink, Richard L. "The Relationship of Self Concept and Selected Personality Variables to Achievement in Music Student Teaching." Journal of Research in Music Education, 3, Fall, 1970.

UNPUBLISHED MATERIALS

Anderson, James Harold. "The Predictive Value of Selected Factors for Student Teacher Success at Western Carolina University." unpublished doctoral dissertation, University of Tennessee, 1968.

Anderson, John Martin. "The Use of Musical Talent, Personality, and Vocational Interest Factors in Predicting Success for Student Music Teachers." unpublished doctoral dissertation, University of Southern California, 1965.

Barth, George William. "Some Personality and Temperament Characteristics of Selected Music Teachers." unpublished doctoral dissertation, University of Southern California, 1961.

Benner, George William. "Some Personality and Temperament Characteristics of Selected Music Teachers." unpublished doctoral dissertation, University of Southern California, 1961

Bodanske, William Holt. "Selected Conditions Associated With The Mobility of Secondary School Music Educators in Missouri." unpublished doctoral dissertation, University of Missouri at Columbia, 1969.

Borkowski, Francis Thomas. "The Relationship of Quality of Work in Undergraduate Music Curricula to Effectiveness of Instrumental Music Teachers in the Public Schools." unpublished doctoral dissertation, West Virginia University, 1967.

Brown, Robert E. "A Study Concerning the Major Weaknesses of Music Teachers in the Secondary Schools of Kansas." unpublished master's thesis, University of Kansas, 1955.

Bullock, Jack A. "An Investigation of the Personal and Professional Attributes of Superior Teachers of Junior High School Instrumental Music in New York State." unpublished doctoral proposal, University of Miami, June, 1972.

Bunyard, Robert Louis. "Factors Affecting the Transfer and Dropout of High School Instrumental Music Teachers." unpublished doctoral dissertation, University of California at Los Angeles, 1963.

Calder, Robert W. "Factors Influencing Male Music Education Graduates of Certain Pennsylvania Institutions of Higher Education to Leave or Not Enter the Profession." unpublished doctoral dissertation, The Pennsylvania State University, 1962.

Dewitt, Charles Jay. "A Study of Selected Variables in Discriminating Between Contrasting Levels of Student Teaching Performance." unpublished doctoral dissertation, University of Virginia, 1970.

Duda, Walter Boleslav. "The Prediction of Three Major Dimensions of Teacher Behavior for Student Teachers in Music Education." unpublished doctoral dissertation, University of Illinois, 1961.

Dvorak, Wayne. "An Investigation of the Colwell Music Achievement Test As An Instrument for Diagnostic Evaluation for Undergraduate Music Education Students." unpublished doctoral dissertation in progress, University of Illinois, 1973.

Edwards, Roger N. "Some Considerations For The Interfacing of the Arts and Evaluation." CEMREL internal paper to be published, March, 1973.

Erbes, Robert Leonard. "The Development of an Observational System for the Analysis of Verbal Interaction in the Rehearsal of Musical Organizations." unpublished doctoral dissertation, University of Illinois, 1972.

Folstrom, Roger James. "A Comparative Study of the Musical Achievement of Students in Three Illinois High Schools." unpublished doctoral dissertation, Northwestern University, 1967.

Fosse, John Benjamin. "The Prediction of Teaching Effectiveness: An Investigation of the Relationships Among High School Band Contest Ratings, Teaching Characterisitcs, and School Environment Factors." unpublished doctoral dissertation, Northwestern University, 1965.

Funk, William Russell. "Comparative Personality Variables, Professional Practices, Opinions and Attitudes in Southern Arizona High School Administrators and Teachers of Performing Groups in Vocal and Instrumental Music." unpublished doctoral dissertation, The University of Arizona, 1973.

George, Warren Edwin. "Significant Predictors for College Achievement in Specified Areas of Music Education and Identification of Potential Graduates (Volumes I and II)." unpublished doctoral dissertation, University of Kansas, 1969.

Gray, John J. "Subject Matter Testing Practices in State University Music Education Curricula." unpublished doctoral dissertation, University of Southern California, 1960.

Krueger, Reynold J. "Extra-Curricular Activities and Music Education Students: A Pilot Study." unpublished research report, Department of Music Education, University of Illinois, January, 1971.

Krueger, Reynold J. "The Oklahoma City University School of Music Testing Program for Freshman Students, 1972-73." unpublished research report, September, 1973.

Lutz, Warren William. "The Personality Characteristics and Experiential Backgrounds of Successful High School Instrumental Music Teachers." unpublished doctoral dissertation, University of Illinois, 1963.

Maier, John Arthur. "An Investigation of the Relationships of Selected Academic and Musical Variables to the Two Year Basic Music Theory Criteria at the State University - Fredonia, New York." unpublished doctoral dissertation, The Ohio State University, 1970.

Morgan, Hazel Nohavec. "The Ability Pattern of Senior College Students Majoring in Music Education at the University of Minnesota." unpublished doctoral dissertation, University of Minnesota, 1946.

Nolin, Wallice H. "Patterns of Teacher-Student Interaction in Selected Junior High School General Music Classes." unpublished doctoral dissertation, The Ohio State University, 1969.

O'Connor, Edward F. Jr. Personal Communication, 1972.

Picerno, Vincent Joseph. "Personal Characteristics of Some Successful Music Teachers in Erie County, New York." unpublished doctoral dissertation, Northwestern University, 1956.

Poland, William. "Measuring Musical Behavior and Predicting Success in the University School of Music." AERA Annual Meeting Presentation, February 14, 1963, Chicago, The Ohio State University.

Rhinehart, Charles B. Jr. "Incidence and Causes of Withdrawal from Public School Music Teaching." unpublished doctoral dissertation, Florida State University, 1963.

Shanks, James Louis. "Concept Achievement in Science and Its Relationship to Some Non-Intellectual Characteristics of Prospective Elementary Teachers." unpublished doctoral dissertation, University of California at Berkeley, 1969.

BEST COPY AVAILABLE

Smith, Martha Lucile. "A Study of Elementary Student Teacher Confidence in and Attitudes Toward Music and Changes that Occur in a Student Teaching Experience." unpublished doctoral dissertation, Michigan State University, 1969.

Sperry, Gale L. "A Study of Music Education Graduates from the University of Minnesota (1947-1956) with Particular Reference to Their Participation and Persistence in the Profession." unpublished doctoral dissertation, University of Minnesota, 1958.

Stake, Robert E. "To Evaluate an Arts Program." unpublished draft copy, University of Illinois at Urbana-Champaign, April 3, 1973.

Strub, Paul. "The Undergraduate Characteristics of Successful Public School Music Teachers." unpublished doctoral dissertation, University of Kansas, 1957.

Turrentine, Edgar M. "Predicting Success in Practice Teaching in Music." unpublished doctoral dissertation, State University of Iowa, 1962.

Watkins, Robert H. "The Construction and Testing of a Forced Choice Scale for the Evaluation of Student Teachers in Vocal Music at the Junior High School Level." unpublished doctoral dissertation, University of Illinois, 1958.

Williams, Raymond Edmund. "The Measurement and Predicting of Cooperating Teacher Effectiveness in Music Education." unpublished doctoral dissertation, University of Illinois, 1958.

Wilson, George Hugh. "A Study of Professional Music Education at Ohio State University." unpublished doctoral dissertation, The Ohio State University, 1958.

APPENDIX A

LIST OF PARTICIPATING SCHOOL DISTRICTS

CALIFORNIA
- Buena Park
- Fountain Valley
- Garden Grove
- Oakland

COLORADO
- Pueblo

FLORIDA
- Tampa
- Gainesville
- Miami

GEORGIA
- Decatur
- Jonesboro

ILLINOIS
- Aurora
- Beecher
- Broadlands
- Bunker Hill
- Champaign
- Carpentersville
- Chebanse
- Chicago
- Cicero
- Clinton
- Cornell
- Country Club Hills
- Danville
- Decatur
- DeLand
- Effingham
- Freeport
- Hillside
- Homewood
- Joliet
- LaGrange
- Lake Zurich
- LeRoy
- Lockport
- Lombard
- McHenry
- Mahomet
- Manhattan
- Mendota
- Momence
- Mt. Pulaski
- Mt. Zion
- Mundelein
- Oak Lawn
- Oglesby
- Peru
- Pittsfield
- Plainfield
- Plymouth
- Princeton
- Princeville
- Quincy
- Roadhouse
- Rockford
- St. Joseph
- Saunemin
- Springfield
- Stockton
- Urbana
- Villa Park
- Westville

- INDIANA
  - Munster
  - South Bend
- KENTUCKY
  - Paducah
- NEW YORK
  - Odessa
- NORTH DAKOTA
  - Fargo
- PUERTO RICO
  - Dorado
- VIRGINIA
  - Norfolk
- WISCONSIN
  - Brookfield
  - Glendale
  - Madison
  - Menomonee Falls
  - Mequon
  - Milwaukee
- IOWA
  - Des Moines
- MISSOURI
  - Rolla
- NORTH CAROLINA
  - Durham
- OHIO
  - Cincinnati
- TEXAS
  - Abilene
- WASHINGTON
  - Tacoma

APPENDIX B

LETTERS

Advisory Board

Frederick K. Fennell
Edwin Gordon
Rose Marie Grentzer
Charles Hoffer
Robert John
William Johnston

George Kyme
Gary Martin
Robert Petzold
Edward Rainbow
Bennett Reimer
Robert Sidnell

Staff

Richard J. Colwell
Roger E. Edwards
Reynold J. Krueger

COUNCIL FOR RESEARCH IN MUSIC EDUCATION

57 EAST ARMORY – UNIVERSITY OF ILLINOIS – CHAMPAIGN, ILLINOIS 61820

TELEPHONE 217-333-1027

28 May 1971

Mr. Patrick Donaghy
Special Projects
Folletts
1010 West Washington Avenue
Chicago, Illinois 60607

Dear Mr. Donaghy:

This letter is in regard to our telephone conversation about the special use of the Music Achievement Tests. Dr. Colwell is my doctoral adviser and has given permission for this project.

Generally, the procedure is to investigate the use of pupil gain scores as a criterion of music teacher accountability. This is within a total purpose of deriving personality profiles for successful music educators. Pupil gain will be one criterion and four types of teacher ratings will also be investigated.

Each of the 300 music educators participating in this study will select 5 music behavioral objectives which they will emphasize in one music class or organization during the fall semester, 1971-72 school year. From these 5 objectives, I will select 3 for objective evaluation with MAT subsections. The sections from the MAT tests which pertain to the selected 3 objectives will be used in a pretest-posttest design to assess the degree to which achievement was successful.

Special test tapes will be made from the MAT records, and machine scored answer sheets will be printed. Correction will be accomplished by the UI Office of Instructional Services, the same office which was used in the piloting stages of the MAT tests.

The use of pupil gain scores as teacher accountability factors is difficult. The technique for this study will be residual gain which accounts for test error and regression effects.

I believe that this will be an exciting investigation. Since over 300 music educators will be participating in this study, many of these persons will be introduced to the MAT tests for the first time. The reason for not using a whole test except in isolated cases where I will be doing a music program evaluation is to make the objective evaluation most appropriate and valid for each individual music educator's situation. I will achieve 1.00 validity because the music teachers must agree to the importance of the objectives which will be evaluated by use of the MAT test subsections.

Cooperatively supported as a service to music education by the Illinois Office of Public Instruction, U. S. Office of Education and the University of Illinois

I appreciate your help and request permission for this special use of the MAT tests. Thank you.

Sincerely,

[S] Reynold J. Krueger

Reynold J. Krueger

FOLLETT EDUCATIONAL CORPORATION

June 1, 1971

Mr. Reynold J. Krueger
Council for Research in Music Education
57 East Armory
University of Illinois
Champaign, Illinois 61820

Dear Mr. Krueger:

You have our permission to use the [illegible] tests as described in your letter of May 28, 1971. Best wishes for what sounds like quite an interesting project.

Sincerely yours,

Patrick Donaghy

Patrick Donaghy
Executive Editor
Special Projects

PD/mo

1010 WEST WASHINGTON BOULEVARD, CHICAGO, ILLINOIS 60607, PHONE 312/666-585

Advisory Board
Frederick K. Fennell
Edwin Gordon
Rose Marie Grentzer
Charles Hoffer
Robert John
William Johnston
George Kyme
Gary Martin
Robert Petzold
Edward Rainbow
Bennett Reimer
Robert Sidnell

Staff
Richard J. Colwell
Roger E. Edwards
Reynold J. Krueger

COUNCIL FOR RESEARCH IN MUSIC EDUCATION
57 EAST ARMORY – UNIVERSITY OF ILLINOIS – CHAMPAIGN, ILLINOIS 61820
TELEPHONE 217–333-1027

Sample of a personal letter sent to music supervisors to enlist their cooperation in the study.

9 August 1971

I am presently drawing the sample for my thesis research study which is investigating personality as a means for predicting music teaching success. This study requires the support of over 250 music educators representing multifarious backgrounds and situations.

I am writing you about the possibility of including your music educators in this study. As a recipient of the Bulletin [and a graduate of Illinois], I know that you are receptive to research. I hope that this project will provide practical and useful information for both public schools and teacher training institutions.

The enclosed materials describe the project in detail. Copies of the tests are included. Generally, I am investigating teacher accountability criteria in terms of pupil growth and teacher ratings. Folletts has granted special permission for using the Colwell Music Achievement Tests so that the pupil gain criterion will be valid for each individual music educator's music teaching situation. The Illinois Teacher Evaluation Questionnaire (ITEQ) will be the rating instrument.

The participating music educators will complete three non-pathological personality inventories. This is the greatest amount of time they will contribute to this study. Each test requires between 30 and 50 minutes, and individual or group testing situations will be arranged.

A request for funding of this project is pending with Region 5 of the U. S. Office of Education. If approved, funds will be available for offering a small gratuity to music educators participating in this study. Hopefully, this monetary incentive will offset volunteer sample bias. Also, funds will be available for data gathering travel. No costs will be incurred by participants or local schools.

I will call you soon in order to discuss this research project. I appreciate your help and look forward to meeting you.

Sincerely,

[S] Reynold J. Krueger

Reynold J. Krueger

Enclosure

Advisory Board
Frederick K. Fennell
Edwin Gordon
Rose Marie Grentzer
Charles Hoffer
Robert John
William Johnston
George Kyme
Gary Martin
Robert Petzold
Edward Rainbow
Bennett Reimer
Robert Sidnell

Staff
Richard J. Colwell
Roger E. Edwards
Reynold J. Krueger

COUNCIL FOR RESEARCH IN MUSIC EDUCATION
57 EAST ARMORY – UNIVERSITY OF ILLINOIS – CHAMPAIGN, ILLINOIS 61820
TELEPHONE 217–333-1027

Dear Music Educator:

Thank you for indicating a willingness to participate in my thesis research study. This investigation requires the active support and enthusiasm of over 300 music educators in a wide variety of music teaching situations in the Midwest, East, and South. I believe that this study is concentrating on some important issues with implications for both public schools and teacher training institutions.

I need your help with this study and would appreciate receiving the packet of materials at your earliest convenience. This will allow time for printing the answer sheets and duplicating test tapes for your pupils as well as attending to numerous administrative requirements.

If you have any questions about this study, please do not hesitate to call me collect. My home phone number is 217-359-2058.

All confidentiality will be maintained.

Sincerely,

[S] Reynold J. Krueger

Reynold J. Krueger

ooperatively supported as a service to music education by the Illinois Office of Public Instruction, U.S. Office of Education and the University of Illinois

BEST COPY AVAILABLE

Advisory Board
Frederick K. Fennell
Edwin Gordon
Rose Marie Grentzer
Charles Hoffer
Robert John
William Johnston
George Kyme
Gary Martin
Robert Petzold
Edward Rainbow
Bennett Reimer
Robert Sidnell

Staff
Richard J. Colwell
Roger E. Edwards
Reynold J. Krueger

COUNCIL FOR RESEARCH IN MUSIC EDUCATION
57 EAST ARMORY – UNIVERSITY OF ILLINOIS – CHAMPAIGN, ILLINOIS 61820
TELEPHONE 217-333-1027

(Sample of the thank-you letter which was individually typed and sent to each music educator's home address upon receipt of completed personality tests--wording varied slightly with each letter)

Dear ____________________:

Thank you for completing the three personality tests and for participating in my thesis research study. I appreciate your help with this exciting investigation.

Soon you will receive the music achievement pretests for administering to one class of students. These will be sent (directly to you at your school address) or (to your music supervisor for distribution to you).

Thank you again for your help with this project, and I wish you success in this new school year.

Sincerely,

[S] Reynold J. Krueger

Reynold J. Krueger

RJK:id

ooperatively supported as a service to music education by the Illinois Office of Public Instruction, U.S. Office of Education and the University of Illinois

Advisory Board

Frederick K. Fennell
Edwin Gordon
Rose Marie Grentzer
Charles Hoffer
Robert John
William Johnston

George Kyme
Gary Martin
Robert Petzold
Edward Rainbow
Bennett Reimer
Robert Sidnell

Staff

Richard J. Colwell
Roger E. Edwards
Reynold J. Krueger

COUNCIL FOR RESEARCH IN MUSIC EDUCATION

57 EAST ARMORY – UNIVERSITY OF ILLINOIS – CHAMPAIGN, ILLINOIS 61820

TELEPHONE 217-333-1027

23 August 1971

Dear Music Educator:

With the 1971-72 school year ready to begin, time is now a crucial factor for my thesis research study which will be conducted during the first semester. Music educators from multifarious situations and backgrounds in the United States are participating in this investigation.

This research study attempts an indepth investigation of the predictive relationships of personality behavioral dimensions to music teaching and student teaching success. I realize that this is an ambitious undertaking, requiring a greater time commitment from participating music educators than previous thesis studies. Nevertheless, one personality investory would not provide reliable and valid information of practical use to public schools and teacher training institutions. The approximate 2½ hours required to complete the three tests are important and necessary.

Defining music teaching success is a most difficult procedure. Previous studies have used a single rating or a raw score pupil gain factor to classify music educators into success groups. However, there is considerable evidence available which cautions against using a single criterion for defining music teaching success. We do not know whether ratings and gain scores are valid music teacher accountability criteria.

Therefore, I decided to include a pupil gain factor and five kinds of ratings in order to determine music teaching success. This study requires that the pupil gain factor be measured by a music achievement test valid for your classroom, thus the reason for your selection of 5 objectives. Special machine scored answer sheets must be printed and test tapes duplicated. We need to pretest one of your classes in September, 1971, and posttest the same class after 16 weeks of instruction. With 250 participating music educators, over 12,500 pupils will be pre and posttested with specially prepared music achievement tests. The preparation of these tests requires at least two weeks.

Please return the packet of test materials to me at your earliest convenience. These packets are expensive and needed for obtaining data from additional music educators.

Naturally, I hope that you will participate in this study. It definitely requires a considerable amount of your time, but this is time well spent.

Cooperatively supported as a service to music education by the Illinois Office of Public Instruction, U.S. Office of Education and the University of Illinois

I appreciate your help with this project and look forward to receiving the packet soon.

Sincerely,

[S] Reynold J. Krueger

Reynold J. Krueger

Music Education Personality Project
57 EAST ARMORY STREET – UNIVERSITY OF ILLINOIS – CHAMPAIGN, ILLINOIS 61820
TELEPHONE (217) 333-1027

Research Committee
Lawrence Aleamoni
Richard Colwell
Charles Leonhard
Maurice Tatsuoka

Director
Reynold J. Krueger

Letter sent personally addressed to each building principal of each participating music educator and music student teacher who was student teaching in his school.

The enclosed prospectus explains in detail my Ed. D. thesis research investigation which is sponsored by the University of Illinois and funded by the U. S. Office of Education through the Regional Research Program. This study includes over 250 music educators representing multifarious teaching situations in the United States as well as 100 student music teachers attending six Illinois teacher training institutions.

Listed below are the music educators in your school or district who have volunteered to participate in this study. I respectfully request your permission to include these music educators in this very exciting project. If you have any questions, please do not hesitate to call me collect. Confidentiality will be maintained and anonymity assured.

Thank you for your help.

Sincerely,

[S] Reynold J. Krueger

Reynold J. Krueger

RJK:id
Enclosure

Participating Music Educator(s):

ooperatively supported by Region V – United States Office of Education and the University of Illinois

Advisory Board
Frederick K. Fennell
Edwin Gordon
Rose Marie Grentzer
Charles Hoffer
Robert John
William Johnston
George Kyme
Gary Martin
Robert Petzold
Edward Rainbow
Bennett Reimer
Robert Sidnell

Staff
Richard J. Colwell
Roger E. Edwards
Reynold J. Krueger

COUNCIL FOR RESEARCH IN MUSIC EDUCATION

57 EAST ARMORY – UNIVERSITY OF ILLINOIS – CHAMPAIGN, ILLINOIS 61820
TELEPHONE 217–333-1027

Personal letter sent to each potential music student teacher and his cooperating teacher.

The enclosed prospectus explains in detail my thesis research study which is investigating the predictive relationship of personality to music teaching and student teaching success. Over 270 music educators in the United States and 100 student music teachers attending six Illinois teacher training institutions are participating in this study which is sponsored by the University of Illinois and funded by the U. S. Office of Education through the Regional Research Program of Region V.

I need your assistance in obtaining two student teacher rating evaluations. These will not affect the student teacher's grade in any way. Rather, these evaluations will be used to investigate the criterion domain of music student teaching success.

Please have one class of your students complete the "Illinois Teacher Evaluation Questionnaire." This will take no longer than ten minutes. This class should be the most mature class with whom the student teacher worked during the student teaching period. This rating evaluation provides the "pupil rating" criterion.

Secondly, please rate the student teacher on the ITEQ form stamped "cooperating teacher." You have spent the most time with the student teacher, and your evaluation is quite important. This study will also investigate "college supervisor," "peer," and "self" ratings.

If you have any questions about this study, please do not hesitate to call me collect or the college supervisor. I appreciate your help with this project.

Sincerely,

[S] Reynold J. Krueger

Reynold J. Krueger

RJK:id
Enclosure

Participating student teacher:

CC: Supervising teacher:

ooperatively supported as a service to music education by the Illinois Office of Public Instruction, U.S. Office of Education and the University of Illinois

Music Education Personality Project

57 EAST ARMORY STREET – UNIVERSITY OF ILLINOIS – CHAMPAIGN, ILLINOIS 61820
TELEPHONE (217) 333-1027

Research Committee
Lawrence Aleamoni
Richard Colwell
Charles Leonhard
Maurice Tatsuoka

Director
Reynold J. Krueger

December 16, 1971

Dear Music Educator:

With this letter, I wish you all the joys of Christmas and a prosperous and successful New Year. Many exciting things have occurred since this research project began, and I want to take this opportunity to bring you up to date.

First, I am most grateful and appreciative of your willingness to participate in this study. In planning this project, I had hopes of obtaining the cooperation of 150 music educators even though some of my colleagues questioned whether that many would really participate. I cannot express enough my sincere thanks to you that nearly 300 of you are giving freely of your time and energy to make this research investigation a success. Music educators are participating from the following states: California, Colorado, Florida, Georgia, Illinois, Inidana, Iowa, Kentucky, Missouri, New Jersey, New York, North Carolina, North Dakota, Ohio, Texas, Virginia, and Wisconsin as well as Puerto Rico.

Second, I was fortunate to have my request for funding of this study approved by Region V of the U. S. Office of Education. While this funding was delayed due to the Wage and Price Freeze, I am pleased that persons in the Government agree with me that this is a most important and needed investigation.

Thank you again for helping with this study. Best wishes for a very Happy Holiday!

Sincerely yours,

[S] Reynold J. Krueger

Reynold J. Krueger
Project Director

RJK:id

Music Education Personality Project

57 EAST ARMORY STREET – UNIVERSITY OF ILLINOIS – CHAMPAIGN, ILLINOIS 61820
TELEPHONE (217) 333-1027

Research Committee
Lawrence Aleamoni
Richard Colwell
Charles Leonhard
Maurice Tatsuoka

Director
Reynold J. Krueger

Dear Music Educator:

Enclosed are the Illinois Teacher Evaluation Questionnaires (ITEQ) for pupil, peer, and self rating evaluations. These should be completed as soon as possible so that the pupil rating results may be returned to you before the end of the school year. A minimum of three weeks is needed for scoring and tabulation.

The enclosed large mailing packet includes ITEQ forms, administration directions, and a class description form for one of your music classes. This class should be the most musically mature and advanced class you teach. This does not need to be the same class used for music achievement testing.

While there should be little difficulty in having secondary school pupils (grades 7-12) complete these evaluations, some elementary pupils may need assistance. If possible, I recommend strongly that another teacher, administrator, counselor, or reliable student administer ITEQ to your pupils. The directions should be read slowly and carefully. If necessary, the test administrator may read each item to the pupils if you anticipate their having difficulty in understanding or reading the items. No more than 15 minutes should be required for completing these pupil evaluations.

The five stamped and addressed peer rating envelopes printed in blue include an amended ITEQ rating blank and an explanatory letter. This form was originally intended for use by pupils so inappropriate items have been crossed out and the wording of some items has been changed.

Please hand these envelopes to five of your fellow teachers who know you well. These peer teachers may teach any subject matter specialty or grade level in your school or attendance area. They should complete their evaluation as soon as possible and mail them directly to me.

The stamped and addressed self rating envelope printed in red includes an amended ITEQ rating blank for you to complete. In order that I may follow up late and lost returns, please list the names and addresses of the peer teachers you requested to cooperate with us in this study on the Peer Teacher List. Include this sheet in the envelope with your self rating.

The following should be returned in the large mailing packet:
1. ITEQ pupil rating blanks 2. Music Class Description Form

If you have any questions about this study or need additional forms, please do not hesitate to call me collect. I certainly appreciate your help with this project. Thank you very much.

Sincerely,

[S] Reynold J. Krueger

Reynold J. Krueger

RJK:id
Enclosures

Cooperatively supported by Region V – United States Office of Education and the University of Illinois

Music Education Personality Project

57 EAST ARMORY STREET – UNIVERSITY OF ILLINOIS – CHAMPAIGN, ILLINOIS 61820
TELEPHONE (217) 333-1027

Research Committee
Lawrence Aleamoni
Richard Colwell
Charles Leonhard
Maurice Tatsuoka

Director
Reynold J. Krueger

(Five copies of this letter were given to each participating music educator for distribution to five peer teachers)

Dear Educator:

A request has been made by one of your school's music educators for you to complete a peer rating which is needed for his or her participation in this research project. The purpose of this study is to investigate the predictive relationship of personality and music teaching success. Over 300 music educators in 16 states and 60 school districts as well as 75 music student teachers attending six Illinois teacher training institutions are participating in this study. This project is sponsored by the University of Illinois and funded by the U. S. Office of Education through the Regional Research Program.

It is extremely difficult to define music teaching success. Therefore the success criteria are under investigation, too, This study is looking at music teacher accountability in terms of pupil growth and four categories of ratings: pupil, administrator, peer, and self.

Enclosed is an amended Illinois Teacher Evaluation Questionnaire (ITEQ). This form was originally designed for use by pupils so inappropriate items have been crossed out and the wording of some items has been changed.

Please fill out the ITEQ rating at your earliest convenience and return it to me in the stamped, self-addressed envelope. Confidentiality will be maintained; anonymity is assured. No ITEQ peer rating results will be returned to the participating music educators.

I sincerely appreciate your help with this project. Thank you very much.

Sincerely,

[S] Reynold J. Krueger

Reynold J. Krueger

RJK:id

Enclosures

Music Education Personality Project

57 EAST ARMORY STREET – UNIVERSITY OF ILLINOIS – CHAMPAIGN, ILLINOIS 61820
TELEPHONE (217) 333-1027

Research Committee
Lawrence Aleamoni
Richard Colwell
Charles Leonhard
Maurice Tatsuoka

Director
Reynold J. Krueger

May 25, 1972

Dear Music Educator:

This research investigation is progressing smoothly although somewhat behind schedule. The rating evaluations have been coming in so well that the office looks like Christmas in May.

To date, 312 music educators and 92 music student teachers are participating. Needless to say, I am delighted and appreciative because a considerable amount of time and effort was required of you.

Do not despair, I will send you your pupil music achievement test scores, pupil rating evaluations, personality test scores, and a synopsis of the findings, conclusions, and recommendations. Due to budgetary considerations, I have had to readjust the schedule of activities. While this investigation has been funded under a grant from the Regional Research Program which limits funding to a maximum of $10,000.00 this project is considerably more expensive and time consuming. I sincerely hope that you will not be too disappointed about receiving feedback later than originally planned.

The personality tests are being scored so that a meaningful interpretation can be provided to you. Music achievement tests are being scored by hand which has delayed this aspect of the study. When all the rating evaluations are received, these will be machine scored in one computer operation.

Your help in making this project successful is most appreciated. Best wishes for a relaxing and enjoyable summer.

Sincerely,

[S] Reynold J. Krueger

Reynold J. Krueger

RJK:rsj

Cooperatively supported by Region V – United States Office of Education and the University of Illinois

Music Education Personality Project

OKLAHOMA CITY UNIVERSITY — SCHOOL OF MUSIC — OKLAHOMA CITY, OKLAHOMA 73106
TELEPHONE (405) 525-5411 - Ext. 2467

Research Committee
Lawrence Aleamoni
Richard Colwell
Charles Leonhard
Maurice Tatsuoka

Director
Reynold J. Krueger

December 9, 1972

Dear Music Educator:

With this letter, I wish you all the joys of Christmas and every happiness in the New Year. This research project is progressing smoothly albeit much behind schedule, and soon you will receive your pupil rating and music achievement test scores.

The enclosed computer output interpretations are from the personality and attitude questionnaires which you took earlier. Consequently, when you look at the printouts, there are a few points you should keep in mind.

There are no such things as good or bad scores on these types of tests. Personality traits that enter into the successful performance of one type of behavior may not help at all in another. For this reason, higher scores do not mean good scores and lower scores do not mean bad scores!! The accuracy of these scores is dependent upon to a large extent how accurately you answered the questions. In a sense, you might say that these scores are reflections of what you are like based on what you said you are like.

All tests are imperfect. The fact that a computer has analyzed the scores does not give tests any greater reliabilities than with which they were designed. If the reports do not match your own evaluations and before you entertain the notion that the tests are right and your own judgment is wrong, please consider the much more likely alternative that the results are simply inaccurate by virtue of their less than perfect reliabilities.

The various statements in the computer printouts are made by comparing your profile of scores with those obtained by psychologists in various research studies. Statistical prediction is not perfect even when the predictions are based on empirically determined relationships.

A considerable amount of good has come for this project to date. I met my fiancé, a participating music educator, through this study! An article has appeared in the Fall issue of the Bulletin of the Council for Research in Music Education, and a report on the preliminary findings will be presented at the North Central MENC meeting in Milwaukee, Wisconsin. With luck, the final report will be available in June, 1973.

Please feel free to write or call me if you have any questions regarding the test interpretations. My home telephone number is (405) 787-9470. Also, please notify me of any changes in your address.

eratively supported by The Regional Research Program — U. S. Office of Education, University of Illinois, and Oklahoma City University

BEST COPY AVAILABLE

Your participation in this project is sincerely appreciated. Thank you very much and have a real nice Holiday!

Cordially,

[S] Reynold J. Krueger

Reynold J. Krueger
Assistant Professor of Music
Director of Music Education

RJK:r
Enclosures

Music Education Personality Project

OKLAHOMA CITY UNIVERSITY – SCHOOL OF MUSIC – OKLAHOMA CITY, OKLAHOMA 73106
TELEPHONE (405) 525-5411 - Ext. 2467

Research Committee
Lawrence Aleamoni
Richard Colwell
Charles Leonhard
Maurice Tatsuoka

Director
Reynold J. Krueger

Music Education Personality Project
Oklahoma City University
School of Music
Oklahoma City, Oklahoma 73106

December 11, 1972

Dear Music Student Teacher:

Enclosed are all of the test interpretations for which I have information for you. If you have not received your ITEQ pupil rating evaluations, it is because I do not have them. Your cooperating teacher did not follow through in obtaining these evaluations from the pupils with whom you worked the best.

Due to the nature of this study, there has been a considerable amount of missing data as far as the music student teacher sample is concerned. The immense amount of mailing required in order to obtain the evaluations necessitated the active cooperation of many persons.

It is important to this project that I have your correct address because in Spring I would like to send you a summary of the findings and results. Please send me a postcard and include your correct address as well as any changes in your name.

By now you are in your new career, and I wish you every success and happiness. Thank you for assisting this project.

Sincerely,

Reynold J. Krueger [S]

Reynold J. Krueger
Assistant Professor of Music
Coordinator of Music Education

RJK:r
Enclosures

peratively supported by The Regional Research Program — U. S. Office of Education, University of Illinois, and Oklahoma City University

Music Education Personality Project

OKLAHOMA CITY UNIVERSITY — SCHOOL OF MUSIC — OKLAHOMA CITY, OKLAHOMA 73106
TELEPHONE (405) 525-5411 - Ext. 2467

Research Committee
Lawrence Aleamoni
Richard Colwell
Charles Leonhard
Maurice Tatsuoka

Director
Reynold J. Krueger

May 14, 1973

Dear Music Educator:

As has been par for the course during this research project, I am much behind schedule in providing you with promised feedback. Please accept my apologies for these delays in communication.

While all the data have been gathered, scored, and analyzed, much additional contemplation is required for interpretation and publication. However, I can share with you some of the preliminary findings to date. Rest assured that upon completion of this project, youwill receive a synopsis of the findings and recommendations as well as your pupils' pre and posttest music achievement scores.

In defining music teaching success for predictive, advisory, and counseling purposes, one must consider that success is at least two dimensional in measurement. This study corroborates previous research findings that ratings are subjective criteria which have little relationships to objective measures such as pupil gain scores.

Music supervisor, building principal, and peer teacher ratings of music educators are significantly interrelated while pupil and music educator self ratings are not. In weighting these rating evaluations by using the first principal component of a factor analysis accounting for 50 percent of the variance in order to arrive at one composite rating score for each music educator, the music supervisor ratings were weighted the most (.83) followed by the building principal (.81), peer teachers (.80), pupil (.59), and music educator self ratings (.39). This indicates that the music supervisors, building principals, and peer teachers were evaluating from similar frames of reference.

While personality temperaments and traits can be accurately measured and defined, these behavioral dimensions should be considered in two theoretical domains: the temperament-trait domain measured by the Sixteen Personality Factor Questionnaire and the dynamic dimension measured by the Motivation Analysis Test. Music teaching success can be predicted from the dynamic variables with pupil learning (gain score) as the criterion. No predictions could be made for music teaching success using the composite rating as the criterion. A person's feelings about himself and others and his drives to succeed and achieve are not only measurable but are significant in predicting music teaching success.

There are distinct personality differences between male and female groups of music educators as there are between subgroups: male instrumental, male choral, female instrumental, and female choral. With more participating music educators in each subgroup, significant predictions probably could be established and cross validated.

Enclosed are your pupil rating evaluations. While I am sure these will be of value to you in your own personal teaching and may provide some very interesting insights about your pupils' perceptions of you; as success criteria, they are of little value in this study.

peratively supported by The Regional Research Program — U. S. Office of Education, University of Illinois, and Oklahoma City University

Your patience and assistance with this project is greatly appreciated, and I wish you an enjoyable and relaxing summer.

Sincerely,

[S] Reynold J. Krueger

Reynold J. Krueger
Assistant Professor of Music
Coordinator of Music Education

RJK:r
Enclosures

APPENDIX C

FORMS USED IN THE STUDY

MUSIC EDUCATOR INFORMATION SHEET

Name__________________ (Mrs. __________) Social Security Number__________
Last, First I. ID Code Number

Home Address____
number street city state zipcode

School____________________Address____________________
____

Home Telephone Number (   )____________School Telephone Number__________

Principal (Mr., Miss, Mrs.)______________________________
____

Music Supervisor______________________________

Number of Years of Teaching Experience through 1972____________________

Highest Degree through 1971__________Credits Beyond Degree______________

Please check teaching duties - You may check more than one:

________General Music
________Band
________Humanities
________Middle or Junior High School
________Choral Music
________Orchestra
________Elementary School
________High School

Please Estimate the Percentage of your Students Studying Music Privately:______%

Choice of Music Objectives - Please fill in the corresponding letter from the Music Objective Selection Sheet:

First Choice____________
Second Choice____________
Third Choice____________
Fourth Choice____________
Fifth Choice____________

MUSIC OBJECTIVE SELECTION SHEET

Please select 5 of the following 15 music behavioral objectives to be included in the learning experiences for the class you selected for Criterion #1 - pupil gain scores on a music achievement test. Please number the objectives from 1 to 5 in order of estimated importance for your class and indicate your selections at the bottom of the INFORMATION SHEET.

From these 5 objectives, 3 will be selected for the pupil evaluation. A specially prepared music achievement test will be compiled from the standardized Music Achievement Test Series by Dr. Richard J. Colwell and published by Folletts. Each objective statement below includes a description of the specific measurement technique used to evaluate the objectives by the Colwell Tests. The Music Achievement Test Series has been developed and evaluated with over 30,000 public school students in the United States since 1967.

Music Objectives

________A.* To differentiate which tone of 2 musical tones is higher or whether the tones are the same pitch. This test consists of 15 items. The pupil answers each item by filling in a blank marked 1, 2, or S for 1st tone higher, 2nd tone higher, or the tones are the same. This is a pitch discrimination objective considered basic to music perception.

________B.* To differentiate which of 3 musical tones heard is the lowest. This is another pitch discrimination objective of musical perception. This test consists of 10 items, each having 3 pitches played on the piano.

________C. To differentiate scalewise intervals from those that leap. This is another perception objective considered basic to acquiring skill in music reading. The test includes 18 items. The pupil decides whether the phrase moves generally in a scalewise manner or generally leaps from one to another ignoring the repeated tones.

________D. To differentiate duple and triple meters - music that moves in twos from music that moves in threes. The pupil listens to a musical phrase of a song played on the piano and decides whether it has 2 or 3 beats per measure. There are 15 test items.

________E. To distinguish major and minor modes in music. The test includes 13 items within a musical situation. Each question contains a phrase of a song, played on the piano, and harmonized appropriately. The pupil decides whether the phrase is entirely in major, entirely in minor, or partly in each mode.

________F.* To determine the key center or key tone of a group of chords in one tonality. This test includes 10 items, each consisting of a standard four chord cadence ending on the tonic chord with the key tone in both soprano and bass. The ability to relate to the key tone is important for intonation and reading activities in performance and for recognition of mode, modulation, and form in listening. After hearing the chords, the student hears 3 single tones. He decides which of the 3, or none, is the key tone.

________ G.* To determine the key center or key tone of a short musical phrase. This test includes 10 items, each consisting of a melodic phrase with harmonic accompaniment. This is similar to #F but is more in keeping within a musical situation.

________ H. To read notes accurately as to pitch. This test includes 12 items. The pupil listens to a 4 measure phrase as he looks at a similar 4 measure phrase notated on his answer sheet. He fills in a blank below every measure in which the notation is different in pitch from the melody he hears. He is not asked to indicate the specific deviation, only the measure in which it occurs. The music is always played correctly, deviations occur in the notation on the pupil's answer sheet.

________ I. To read notes accurately as to rhythm. This test includes 14 items. The construction is identical to #H, except that errors of rhythm rather than pitch occur. The pupil fills in a blank below every measure in which the notation is rhythmically different from the melody he hears. He makes 4 discriminations for each item.

________ J. To recognize whether two chords are the same or different and where different, to find which note is altered. The test consists of 20 items. The first 4 note chord is presented in block form, immediately followed by an arpeggiated chord containing pitches identical to the first chord, or with a single pitch altered. The arpeggiated version is used for the second chord to ensure that the pupil is able to identify correctly the voice in which a change may have occurred. Many music programs emphasize the importance of chord structure through singing, hearing, and writing activities. In performing groups, the hope exists that use of such exercises will transfer to musical situations in the form of better intonation, balance, and more skillful treatment of phrases and cadences.

________ K. To recognize the melody and determine whether it appears in the top, middle, or bottom voice of a harmonized version. This task presents a practical musical problem, for when students listen to music, they are aware that the work is made up of different lines, different parts, different voices; and that to make sense of the composition, the important lines must be distinguished from those less important. The test consists of 20 items. The melody is presented first on the piano, then repeated in a three part setting played by violin (top voice), viola (middle voice), and cello (lowest voice). In the three part setting, the melody always appears in the same octave as that in which it is originally introduced.

________ L. To identify the instruments of the band and orchestra by sound. This test consists of 15 items which measure the student's ability to select the correct instrument from four possible choices, with a fifth being that none of the four listed is correct. This duplicates a commonly used school approach where instruments are demonstrated individually, live or on record. 10 items have the instrument played alone while 5 items measure the student's ability to correctly identify instrumental timbre in a solo context within an orchestral setting.

_________M. To identify music of different styles by selecting representative composers. This test consists of 20 items, each including a short orchestral excerpt. The student selects from four choices the composer whose style most closely resembles that of the musical excerpt. The foils all represent stylistic periods differing from the correct answer.

_________N. To differentiate monophonic, homophonic, and polyphonic music. These represent the basic three approaches to musical compositions. This test includes 20 items, each consisting of a musical excerpt of sufficient length to give evidence of its texture. These are performed on the piano.

_________O. To differentiate half, full, and deceptive cadences. This test includes 15 items, each consisting of a short musical phrase ending with a cadence. The item is played on the piano and is of sufficient length to establish tonality. A basic chord function is the cadence which creates a pause of varying degree and helps to define the phrase, as well as establishing tonality and modulation. The cadences used are: half - I - V; Full - V - I; and deceptive - V - VI.

Each of the test sections includes two or more musical examples for the pupil to answer. Directions are included on the test, and teachers may explain the test sections in more detail if desired.

No music achievement test can possibly evaluate all the musical learnings of elementary and secondary school pupils. The music objectives evaluated by the Music Achievement Test Series are those which have consistently been considered important by leading music educators and are included in most methods books.

The purpose of having each music educator select 5 objectives is to ensure the validity of the music achievement test for use in the music learning situation. While there may be many other objectives achieved in a year's time of instruction which are quite important and are not evaluated by the Music Achievement Test Series, it is hoped that some will be able to be objectively evaluated.

Each teacher will administer the test to the selected class. The investigator will have the answer sheets scored and will mail results to each teacher for both test administrations. The use of pupil gain scores has received considerable emphasis today in educational circles. Yet, these are very elusive criteria of teaching and evaluation. The statistical technique used in this study will not compare groups of pupils. Instead, each teacher's class will be held constant. All the factors of pupil differences and achievement levels are accounted for in the residual gain technique which will be employed in this study. It is hoped that this investigation will be able to answer some of the probing questions concerning pupil gain scores, educational contracting, and the like.

* Tests A and B and F and G were later combined into two tests, AB and FG by the investigator. There were too few items in tests B and G to be used as separate test sections.

STUDENT TEACHER INFORMATION SHEET

Name______________________________________University____________________
(last) (first) (i)

Student Teaching Center(s)____

Cooperating Teacher(s)____

School Address____

Student Teacher's Home Address____
(for sending project results and test scores)

University or College Music Supervisor__________________________________

Adviser______________________________

Period for Student Teaching_____________________to______________________
(beginning date) (ending date)

Return this sheet and tests to:

Music Education Personality Project
57 East Armory Street
Champaign, Illinois 61820

BEST COPY AVAILABLE

MUSIC ACHIEVEMENT TEST (MAT) INFORMATION SHEET

(Please fill out this sheet and return it with the MAT answer sheets and test tape)

1. Music educator's name____________________ 2. MAT test sections____________

3. Date test was given__________________________

4. Type of music class tested____

5. Number of times per week this class meets for music instruction.____________

6. Number of minutes this class meets for each session_______________________

7. Date when your second semester begins___________________________________

8. Please write comments or questions below:

MUSIC ACHIEVEMENT TEST
Posttest Information Sheet

(Please fill out this sheet and return it with the MAT answer sheets and te[illegible])

1. Music educator's name____

2. School name and address____

____

3. MAT test sections___________ 4. Date posttest was given________________

5. Comments:

MUSIC CLASS DESCRIPTION FORM
(ITEQ)

Please fill in the requested information and return with the ITEQ sheets. Results will be mailed to you upon conclusion of this study.

1. Music educator's name__________________________________

2. Date ITEQ was administered______________________________

3. Type of music class____________________________________

4. Number of times per week this class meets for music instruction_____________________________

5. Number of minutes this class meets for each session____

6. Who administered ITEQ? (check one)

    _____a. building principal

    _____b. another teacher

    _____c. counselor

    _____d. student

    _____e. music educator named above

    _____f. _________________________

7. Please write comments or questions below:

ILLINOIS TEACHER EVALUATION QUESTIONNAIRE

Class Description Form

Student Teacher

Cooperating teacher:
Please fill in the requested information and return with ITEQ sheets.

1. Music student teacher's name______________________________

2. Cooperating teacher's name______________________________

3. College or University______________________________

4. Date ITEQ was administered______________________________

5. Type of music class______________________________

6. Number of times per week this class meets for music instruction________

7. Number of minutes this class meets for music instruction__________

8. For how many weeks did this music student teacher work with you?______

9. If you would like to receive a synopsis of the findings and conclusions of this research study, please write your home address below:

____
(number) (street)

____
(city) (state) (zipcode)

10. Please write comments or questions below:

ILLINOIS TEACHER EVALUATION QUESTIONNAIRE

Administrator Rating For__________________________________

School__________________________________

Please fill in the requested information and return with the ITEQ form. All confidentiality will be maintained: anonymity is assured. No ITEQ administrator rating results will be returned to the participating music educator.

A synopsis of the findings and conclusions of this study will be mailed to all administrators. Copies of the final report will be available to school systems upon request.

1. Your name____
2. Position:_______Building Principal or_______Music Supervisor
3. Number of years in present position____________________________
4. number of years in school administration__________________________
5. Number of years of school teaching experience________________________
6. Subjects taught____
7. Number of years this music educator has been under your supervision_____
8. Please check the category which best describes your school:

   _______a. Socially deprived
   This school includes a majority of culturally or socially deprived pupils and a substantial number of Title I pupils whose parents are near poverty income levels. Pupil reading levels would be below nationally established grade norms. These schools are found typically in the inner city areas of large urban communities.

   _______b. Socially advantaged
   This school includes a majority of pupils whose parents are earning middle class or higher incomes. Relatively few Title I pupils attend this school. A socially advantaged school is found typically in tye suburbs, average size cities, and fringe areas of very large urban communities.

9. Please write comments or questions below:

PEER TEACHER LIST

1. ____
First Name Initial Last Name

____
Street Address City State Zip

2. ____
First Name Initial Last Name

____
Street Address City State Zip

3. ____
First Name Initial Last Name

____
Street Address City State Zip

4. ____
First Name Initial Last Name

____
Street Address City State Zip

5. ____
First Name Initial Last Name

____
Street Address City State Zip

RETURN IN THE SELF RATING ENVELOPE!

[illegible] FORM [illegible]

Measurement and Research Division, Office of Instructional Resources, UNIVERSITY OF ILLINOIS © By Richard E. Spencer, 1968

DIRECTIONS: 1. PRINT INSTRUCTOR'S LAST NAME HERE ____________________
2. COMPLETE IDENTIFICATION INFORMATION BELOW AS THE TEACHER INDICATES.
3. RESPOND TO THE ITEMS PRESENTED FRANKLY AND COMPLETELY - ONE RESPONSE PER ITEM.
4. USE PENCIL ONLY - DO NOT USE PEN, BALL POINT OR INK OF ANY KIND.

422
(SEE SAMPLE MARK AND RESPONSE CODE)

MAKE ERASURES
AS COMPLETE
AS POSSIBLE

6 ORDER

5

4 EXPECTED GRADE IN COURSE

3 DO YOU [illegible] YES NO

7 AGRICULTURE ART BUS. ED. ENGLISH FOR. LANG. HEALTH HOME EC. IND. ED. MATH MUSIC PE MEN PE WOMEN SCIENCE SOC. STU. SPEECH OTHER

9 SEX

8 TODAY'S DATE DAY YEAR

1 The teacher seems confident and knowledgeable.
2 The class can easily deceive the teacher.
3 The teacher knows what she (he) is talking about.
4 I like this teacher.
5 This teacher has 'favorites' in class.
6 This subject is quite interesting.
7 Quite a boring teacher.
8 Discipline is a problem in this class.
9 This subject is easy to follow.
10 The teacher considers teaching a chore.
11 The teacher seems to talk down to students.
12 The class material is poorly organized.
13 This class is not much fun.
14 The teacher seems to feel afraid she (he) will do something wrong.
15 The teacher gives extra attention to students willingly.
16 I think this teacher teaches quite well.
17 Out of all the teachers I have had, this teacher ranks quite high.
18 The pace of teaching is too slow.
19 The teacher seems confused.
20 The tests are too difficult.
21 Generally, the teacher seems well organized.
22 The teacher changes the method of teaching to suit the occasion.
23 An exciting teacher.
24 I would like to have another teacher like this one.
25 This class is a waste of time.
26 Overall, this subject is one of my favorites.
27 The text material is very good.
28 More teachers should be like this.
29 This class is quite rewarding.
30 Not much is gained by studying this subject.
31 The teacher encourages new ideas and viewpoints.
32 Homework assignments are helpful in understanding the subject.
33 Not enough student discussion.
34 The teacher has a thorough knowledge of the subject.
35 The types of test questions used in this class are good.
36 The teacher does not explain things very well.
37 An uninteresting teacher.
38 This class is one of the poorest of all my classes.
39 The teacher gets off the point too often.
40 Many other teachers I have had can teach better.

SAMPLE MARKS:

USE
PENCIL
ONLY

a) A ▌
b) A ▌
c) ▌ A
d) A ▌

RESPONSE CODE:

MARK A IF YOU STRONGLY AGREE WITH THE ITEM
MARK B IF YOU AGREE MODERATELY WITH THE ITEM
MARK C IF YOU DISAGREE MODERATELY WITH THE ITEM
MARK D IF YOU STRONGLY DISAGREE WITH THE ITEM

ILLINOIS TEACHER EVALUATION QUESTIONNAIRE — FORM 66

Measurement and Research Division, Office of Instructional Resources, UNIVERSITY OF ILLINOIS © By Richard E. Spencer, 1968

DIRECTIONS
1. PRINT INSTRUCTOR'S LAST NAME HERE ______________
2. COMPLETE IDENTIFICATION INFORMATION BELOW AS THE TEACHER INDICATES.
3. RESPOND TO THE ITEMS PRESENTED FRANKLY AND COMPLETELY—ONE RESPONSE PER ITEM. (SEE SAMPLE MARK AND RESPONSE CODE)
4. USE PENCIL ONLY—DO NOT USE PEN, BALL POINT OR INK OF ANY KIND.

423

MAKE ERASURES AS COMPLETE AS POSSIBLE

3 DO YOU PLAN TO GO TO COLLEGE YES NO

4 EXPECTED GRADE IN THIS COURSE

5

6 GRADE

7 AGRICULTURE ART BUS. ED. ENGLISH FOR. LANG. HEALTH HOME EC. IND. ED. MATH MUSIC PE MEN PE WOMEN SCIENCE SOC. STU. SPEECH OTHER

8 TODAY'S DATE DAY YEAR

9 SEX

1 The teacher seems confident and knowledgeable.

2 ~~[illegible]~~

3 The teacher knows what she (he) is talking about.

4 I like this teacher.

5 This teacher has "favorites" in class.

6 ~~[illegible]~~

7 Quite a boring teacher.

8 Discipline is a problem ~~[illegible]~~ WITH HIS (HER) CLASSES.

9 ~~[illegible]~~

10 The teacher considers teaching a chore.

11 The teacher seems to talk down to students.

12 ~~[illegible]~~

13 ~~[illegible]~~

14 The teacher seems to be afraid she (he) will do something wrong.

15 The teacher gives extra attention to students willingly.

16 I think this teacher teaches quite well.

17 Out of all the teachers I ~~have had, this teacher ranks quite high.~~ KNOW, THIS TEACHER RANKS QUITE HIGH.

18 ~~[illegible]~~

19 The teacher seems confused.

20 ~~[illegible]~~

21 Generally, the teacher seems well organized.

22 The teacher changes the method of teaching to suit the occasion.

23 An exciting teacher.

24 I would like to have another teacher like this one, ON MY STAFF OR TEACHING AT THIS SCHOOL.

25 ~~[illegible]~~

26 ~~[illegible]~~

27 ~~The text [illegible] is very good.~~

28 More teachers should be like this.

29 ~~[illegible]~~

30 ~~[illegible]~~

31 The teacher encourages new ideas and viewpoints.

32 ~~[illegible]~~

33 ~~[illegible]~~

34 The teacher has a thorough knowledge of ~~the subject~~ MUSIC.

35 ~~[illegible]~~

36 The teacher does not explain things very well.

37 An uninteresting teacher.

38 ~~[illegible]~~

39 ~~[illegible]~~

40 Many other teachers I ~~have had can teach better.~~ KNOW CAN TEACH BETTER.

SAMPLE MARKS:

USE PENCIL ONLY

a) ▮

b) ▮

c) ▮

d) ▮

RESPONSE CODE:

MARK IF YOU STRONGLY AGREE WITH THE ITEM

MARK IF YOU AGREE MODERATELY WITH THE ITEM

MARK IF YOU DISAGREE MODERATELY WITH THE ITEM

MARK IF YOU STRONGLY DISAGREE WITH THE ITEM

VITA

Reynold Jack Krueger was born 26 June 1942 in Milwaukee, Wisconsin, where he attended public elementary and secondary schools. He graduated from the University of Wisconsin at Madison with a Bachelor of Music degree in instrumental music education in 1964, a Master of Science degree in choral music education in 1967, and from the University of Illinois at Urbana-Champaign, a Doctor of Education degree in music education, administration, and supervision. Presently, he is Assistant Professor of Music and Coordinator of Music Education at Oklahoma City University.

Professor Krueger has five years of contractual public elementary and secondary school music teaching experience in choral, general, and instrumental music. He has been a manager of the University of Wisconsin Bands; a research, teaching, and editorial assistant at the University of Illinois; and a supervisor of music student teachers as well as a speaker and participant at MENC regional and national conventions. He is an active conductor and adjudicator and is a member of the Council for Research in Music Education, Society for Research in Music Education, Music Educators National Conference, Oklahoma Music Educators Association, Sigma Phi Epsilon, Phi Delta Kappa, Lions International, and faculty adviser Delta Zeta Chapter of Phi Mu Alpha Sinfonia.